SELECTION INDEX AND INTRODUCTION TO MIXED MODEL METHODS

SELECTION INDEX

CRC Press
Boca Raton Ann Arbor London Tokyo

Library of Congress Cataloging-in-Publication Data

Van Vleck, L. Dale (Lloyd Dale), 1933–
Selection index and introduction to mixed model methods for genetic improvement of animals : the green book / L. Dale Van Vleck.
p. cm.
Includes bibliographical references.
ISBN 0-8493-8762-0
1. Selection indexes (Animal breeding)—Statistical methods.
I. Title. II. Title: Green book.
SF105.26.V36 1993
636.08'21—dc20 93-8547
CIP

Direct all inquiries to CRC Press, Inc., 2000 Corporate Blvd., N.W., Boca Raton, Florida 33431.

International Standard Book Number 0-8493-8762-0

Printed in the United States of America 1 2 3 4 5 6 7 8 9 0

Printed on acid-free paper

University of Nebraska Lincoln
Institute of Agriculture and Natural Resources

Animal Science Department
P.O. Box 830908
Lincoln, NE 68583-0908
FAX (402) 472-6362

July 16, 1999

TO: Elana Meyers, CRC Press

FROM: Dale Van Vleck Dale Van Vleck

I request permission to copy for teaching purposes the text book of which I am the author but which is out of print:

Selection Index and Introduction to Mixed Model Methods

Your catalog number: BW8762 ISBN: 0849387620

My fax number is 402/472-6362.

Thank you.

FAXED 7/16/99

CRC PRESS, LLC
2000 Corporate Blvd. N.W.
Boca Raton, FL
33431

PERMISSION TO REPRINT IS GRANTED BY CRC PRESS, LLC
CREDIT LINE REQUIRED: Reprinted with permission from (full book or journal reference). Copyright CRC Press, Boca Raton, Florida

7/16/99 Date — Permissions Manager

Permission is granted, subject to the following terms and conditions:

Permission is granted for one-time reproduction of the specified material only for the purposes designated in your original request and does not extend to future editions of your Work;

Permission applies to Print format only and does not extend to Electronic customized versions, derivatives or ancillary works. Permission for such must be separately requested.

A fee of N/C shall be paid for use of copyright materials.

Permission must be obtained from original copyright holder for materials in our work that credits another source.

Special conditions/exceptions: See attached Author Letter

THE AUTHOR

L. Dale Van Vleck is a Research Geneticist with the Agricultural Research Service, USDA, and Professor of Animal Science at the University of Nebraska, Lincoln. For 31 years, he was on the faculty of Cornell University where he taught quantitative genetics and led the research effort in genetic evaluation methods.

He also has taught at the University of California, Davis and at the Scandinavian Graduate Course at the Swedish Agricultural University in Uppsala.

Dr. Van Vleck received the B.S. degree fron UNL with high distinction and his M.S. one year later. He obtained his Ph.D. from Cornell University in Animal Breeding with minors in statistics and physiology. In 1986, he was awarded an honorary Doctor of Science degree from the University of Nebraska.

Dr. Van Vleck is a member of the American Society of Animal Science, the American Dairy Science Association, the American Genetic Association, the Biometric Society and is an editor for Theoretical and Applied Genetics, on the editorial board of the Brazilian Journal of Genetics, and the international editorial advisory board of Genetics, Selection, Evolution.

Dr. Van Vleck was an Alpha Zeta Foundation Fellow and a National Science Foundation Fellow during his graduate programs. He is a member of the honorary societies Sigma Xi, Alpha Zeta, Gamma Sigma Delta and Phi Kappa Phi. His national awards include the Animal Breeding and Genetics award of the American Society of Animal Science, the National Association of Animal Breeders award in Genetics and the Jay L. Lush award in Animal Breeding and Genetics of the American Dairy Science Association.

Dr. Van Vleck has given many invited lectures at international and national meetings and universities. He has published more than 200 papers, 126 abstracts, and 140 popular articles on genetics of animals. He is an author or editor of six books and numerous chapters in books. His current research centers on developing methods for genetic evaluation and selection of meat animals.

. . . to students of genetic selection everywhere,

and to,

C. R. Henderson, who led us to

better ways . . . .

CONTENTS

PART II. INTRODUCTION TO MIXED MODEL PREDICTION

PART ONE

SELECTION INDEX

The following 22 chapters cover material needed for application of the selection index to nearly any imaginable selection problem. The initial chapters review or introduce expected values, probabilities for identity by descent, kinds of genetic values and genetic variances, and genetic covariances between relatives. These tools and definitions provide the foundation for deriving and applying the selection index and its properties. Much of this material was developed by Dr. C. R. Henderson who began such a course in 1948 at Cornell University after studying at Iowa State University with Dr. Jay L. Lush and Dr. L. N. Hazel. For many years, I taught the course, also at Cornell University, with some new material based usually on the principles outlined by Henderson. The course has subsequently been given at the University of Nebraska-Lincoln.

CHAPTER 1

PARAMETERS, STATISTICS, AND EXPECTED VALUES

A review of some basic statistics may be useful before discussing selection for quantitative traits. Two important parameters for the description of continuous or quantitative traits are the mean (or average) and the standard deviation. The usual Greek symbols for these are μ, "mu," and σ, small "sigma." The square of the standard deviation, σ^2, is called the variance, "sigma squared."

A subscript can be used to distinguish the means and variances for different populations or different traits; e.g., μ_X and σ_X^2 would be the mean and variance for some trait called X. After this chapter when developing selection index procedures, means, variances, and covariances will be assumed to be known exactly. When true values are known exactly, they are called parameters. In the real world, parameters are never known exactly because the entire potential population is not known or measured. Then parameters such as the mean or variance must be estimated from a sample of the population. Technically, estimates are known as statistics. A statistic that estimates a parameter may be given the same symbol but with a hat or caret, ^, to distinguish it from the parameter. Often parameters are designated with Greek letters and estimates by corresponding English letters in some way. Pretending that parameters are known can be justified in some cases because with large samples the error (difference of estimate from parameter) will be very

small. In the last part of the book on mixed model procedures, this pretence will be tightened so that means are estimated from the population jointly with predicting genetic values. Even then, variances and covariances will be assumed to be known exactly. Chapters 35-39 describe some simple ways of estimating variances and covariances.

THE MEAN

If x_i ($i = 1, ..., n$) is the observation on the i^{th} individual for trait X, then the estimate of μ_x is $\hat{\mu}_x$ or equivalently $\bar{x}$, "x-bar". A simple estimate is

$$\hat{\mu}_x = \sum_{i=1}^{n} x_i/n = (x_1 + x_2 + \cdots + x_n)/n \text{ which may also be called } \bar{x}_n,$$

the average of n observations. The symbol $\sum_{i=1}^{n}$ is mathematical notation that means to sum everything that follows the Σ for changes in the subscript i which changes by units of 1 from $i = 1$ (the first record) to $i = n$ (the last record). If observations for the whole population, N, were known, then $\mu_x = \sum_{i=1}^{N} x_i/N$.

VARIANCE

Although the standard deviation is a more intuitive measure of variability, the usual measure of variability is the variance, σ_x^2, which is the standard deviation squared for trait X. Knowledge of variances is necessary in animal breeding for at least two reasons. Variances are useful in describing populations and, more importantly, are used along with covariances in developing procedures for predicting genetic values. The definition of σ_x^2 for a population is

$$\sigma_x^2 = E[(x_i-\mu_x)^2] = [(x_1-\mu_x)^2 + (x_2-\mu_x)^2 + \ldots + (x_N-\mu_x)^2]/N,$$

where N is the total number of observations in the population. The E stands for expected or average value and will be discussed later in this chapter. Thus, σ_X^2 is a kind of average, i.e., is the average of the squared deviations of the observations from the mean for a variable named X. While the variance is in terms of units squared, the standard deviation is in terms of the units of measurement--the same as the mean, e.g., the mean of milk production may be expressed in lb. of milk, the variance in lb^2 of milk, and the standard deviation in lb. of milk.

Computing the Variance

If x_i (i=1, ..., N) is the observation on the i^{th} individual, then

$$\sigma_X^2 = \sum_{i=1}^{N} (x_i-\mu_X)^2/N = (\Sigma x_i^2-N\mu_X^2)/N.$$

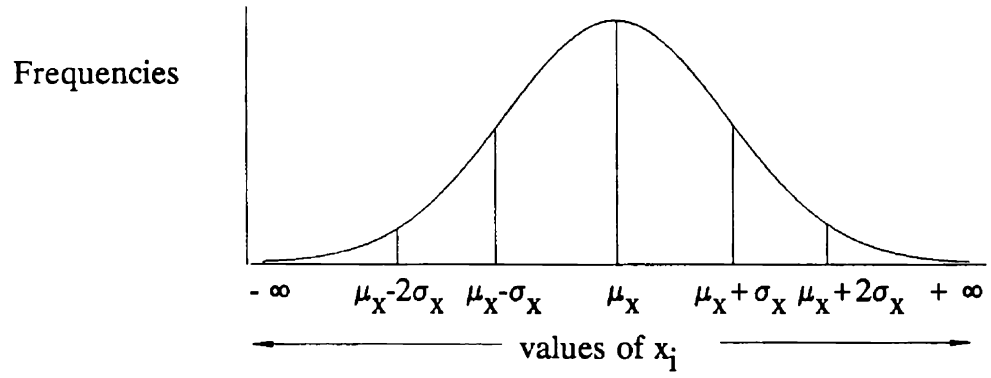
The above procedure is appropriate when μ_X is known exactly. When N includes the whole population, the computed value is the population variance, and when n is a sample of the population, the value is an estimate of the population variance and should be denoted as $\hat{\sigma}_X^2$. If μ_X is estimated from a sample of the data as $\bar{x}$ or $\hat{\mu}_X$, then σ_X^2 is estimated as:

$$\hat{\sigma}_X^2 = [\Sigma x_i^2 - \frac{(\Sigma x)(\Sigma x)}{n}]/(n-1),$$

where n-1 is the degrees of freedom.

The division is by n-1 so that $E(\hat{\sigma}_X^2) = \sigma_X^2$, $= \sigma_X^2$, i.e., the average of estimates of σ_X^2 will be σ_X^2. Thus, the estimate is said to be unbiased. Alternative computing procedures are listed in Table 1.1. The following section on expected values will describe how to find the expected or average value of estimates such as $\hat{\sigma}_X^2$.

The mean and standard deviation characterize a normal distribution of observations. The normal distribution follows the bell-shaped curve where the values along the horizontal axis are plotted against the frequencies of those values on the vertical axis.



The average of all the x_i is μ_X and lies at the center of the symmetrical distribution--one-half the x_i above and one-half the x_i below μ_X. The range $\mu_X - \sigma_X$ to $\mu_X + \sigma_X$ will contain 68% of the x_i; $\mu_X - 2\sigma_X$ to $\mu_X + 2\sigma_X$ will contain 96% of the x_i. Multipliers of σ_X for other frequencies are given in most introductory statistics books.

The distribution of averages of n observations has mean μ_X. The variance of the averages is $\hat{\sigma}^2_X/n$, with the square root, $\hat{\sigma}_{\bar{x}} = \hat{\sigma}_X/\sqrt{n}$, called the standard error of the mean.

COVARIANCE

The variance and the corresponding standard deviation thus measure how one trait varies. The covariance, a measure of how two traits vary together (co-vary), is also needed in developing selection procedures. For example, the covariance between two traits measured on the same animal, e.g., height and weight, may be needed or the covariance between the same trait measured on two relatives may be needed. The definition and computing procedures for the covariance are analogous to those for the variance.

Suppose that the two measures are x_i and y_i ($i=1, ..., N$) for the measurements on the i^{th} pair of relatives. The covariance has the symbol σ_{xy} (sigma-x-y) and is defined as the average of products of deviations from the means of traits X and Y;

$$\sigma_{xy} = E[(x_i-\mu_x)(y_i-\mu_y)]$$

$$= [(x_1-\mu_x)(y_1-\mu_y) + (x_2-\mu_x)(y_2-\mu_y) + \ldots + (x_N-\mu_x)(y_N-\mu_y)]/N$$

$$= \sum_{i=1}^{N} [(x_i-\mu_x)(y_i-\mu_y)]/N.$$

This computation is appropriate when μ_x and μ_y are known exactly and gives the population covariance when N includes the whole population and gives an estimate ($\hat{\sigma}_{xy}$) of the population covariance when n is a sample.

If μ_x and μ_y are estimated from a sample of the population as $\bar{x}$ and $\bar{y}$, ($\hat{\mu}_x$, $\hat{\mu}_y$), then

$$\hat{\sigma}_{xy} = \Sigma[(x_i-\hat{\mu}_x)(y_i-\hat{\mu}_y)]/(n-1)$$

$$= [\Sigma x_i y_i - \frac{(\Sigma x_i)(\Sigma y_i)}{n}]/(n-1).$$

Note the similarity of the computing procedures for variances and covariances.

A positive covariance indicates that as the value for one trait increases, the value for the other trait also tends to increase. A negative covariance indicates that as the value for one trait increases, the other tends to decrease. The traits are not correlated when the covariance is zero. The units of a covariance are units of the first trait times units of the second trait.

FREQUENCY DATA

Sometimes observations fall into discrete categories. An example is with two alleles at one locus which results in three genotypes--the two homozygotes and the heterozygote. Estimates of means, variances and covariances can be made with the previously described procedures where each observation is treated separately. For purposes of algebraic simplification or slightly less tedious computation, advantage can be taken of all observations in a category being the same. Suppose that the observations in c categories are y_i (i=1, ..., c) with frequencies in each category of f_i. Unless some categories are thrown out, $\sum_{i=1}^{c} f_i = 1$. The formulas are given for cases where Σf_i might not equal unity. Suppose that the sample size is n. Then the number with observation y_i in category i is nf_i. The sum of all observations is $\sum_{i=1}^{c} nf_i y_i$. The estimated mean is the sum divided by the number of observations, $\sum_{i=1}^{c} nf_i$;

$$\hat{\mu}_y = \bar{y} = (\Sigma nf_i y_i)/(\Sigma nf_i) = \Sigma f_i y_i / \Sigma f_i \, .$$

The definition of variance is the same as for non-categorized data but the squared deviations from the mean are identical for all observations in the same category, i.e., nf_i will be $(y_i - \hat{\mu}_y)^2$ so that the sum of squared deviations is

$$\Sigma nf_i(y_i - \hat{\mu}_y)^2 = n\,\Sigma f_i(y_i - \hat{\mu}_y)^2 .$$

Division by number of observations minus one results in:

$$\hat{\sigma}_y^2 = [n\Sigma f_i(y_i - \hat{\mu}_y)^2]/(n\Sigma f_i - 1) \quad \text{and for } \Sigma f_i = 1$$

$$\hat{\sigma}_y^2 = [\Sigma f_i(y_i - \hat{\mu}_y)^2]\,[n/(n-1)] .$$

The preceding formulas are given in Table 1.1. Algebraically identical computing formulas which in most cases are easier to use are in the right column of Table 1.1.

CORRELATION

The correlation coefficient is a standardized measure of the relationship between two traits which allows comparisons of correlations among different pairs of traits. The possible range is -1 to +1 with no units involved for either trait. The correlation between traits X and Y or relatives x and y is defined as

$$r_{xy} = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2 \sigma_y^2}} .$$

If estimates of the covariance and variances are used in the formula, then r_{xy} is an estimate of the population correlation coefficient.

REGRESSION

The selection index procedures that will be developed are related to regression equations. The simplest form of a regression equation is to predict the value of some trait for an animal i, e.g., y_i, when the measurement of another trait, x_i, is known. The statistical procedure of minimizing the squared error between y_i and the prediction, $\hat{y}_i$, results in the equation for regression of trait Y on trait X. The basic principle is that a change in trait X results in a corresponding change in trait Y. The regression coefficient denoted as $b_{y \cdot x}$ (b-y-dot-x), describes the magnitude of the corresponding change. Least squares or minimization of $\Sigma(y_i - \hat{y}_i)^2$ results in the solution for $b_{y \cdot x} = \hat{\sigma}_{yx} / \hat{\sigma}_x^2$. The full equation also depends on $\hat{\mu}_y$ and $\hat{\mu}_x$:

$$\hat{y}_i = \hat{\mu}_y + b_{y \cdot x}(x_i - \hat{\mu}_x) .$$

As with the correlation coefficient, the covariance determines the direction of the change. Note that the covariance is in terms of units for x by units for y and the variance of X is in

the square of units for X so that $b_{y \cdot x}$ is in terms of units of Y divided by units of X, i.e., change in Y per unit change in X. The estimated means $\hat{\mu}_y$ and $\hat{\mu}_x$ are as described earlier. To estimate μ_y, μ_x and $b_{y \cdot x}$, measurements on both traits are needed. Then, to predict y_i as $\hat{y}_i$, only a measure on x_i is needed.

EXPECTED VALUES

The use of expected values increases the powerful and flexibility of the selection index but at the expense of minor frustration of some students who initially have difficulty in developing a feeling for what they are doing. Experience has shown that most students overcome this difficulty after some practice and that they become much more adept at solving problems which involve more than the usual case of selection for additive genetic value.

The symbol often used for the expected or average value of some expression involving constants and variables is E( ). Expected values of most expressions used in estimating genetic parameters are relatively easy to determine if six definitions are remembered.

Let c = constant; x_i = variable from some distribution with mean μ_x and variance σ_x^2; and y_i = variable from some distribution with mean μ_y, variance σ_y^2, and covariance with x_i, σ_{xy}.

Definition 1: $E(c) = c$. Certainly the average value of a constant is that constant. Similarly $E(c^2) = c^2$.

Definition 2: $E(x_i) = \mu_x$. The average of all possible values of variable X is its average or mean, μ_x.

Definition 3: $E(cx_i) = c\,E(x_i) = c\mu_x$. The average of all possible values of a variable times a constant is the constant times the mean of the variable. The principle is that for expressions involving a constant the constant can be taken outside the expectation operation.

Definition 4: $E(x_i+y_i) = E(x_i) + E(y_i) = \mu_x + \mu_y$. The principle is that the expectation of a sum can be taken as the sum of the expectations of the parts.

Definition 5: $E[(x_i-\mu_x)^2] = \sigma_x^2$. By definition, the variance of a variable X, σ_x^2, is the average squared deviation of the variable from its mean. Definition 5 leads directly to $E(x_i^2) = \sigma_x^2 + \mu_x^2$. If the equation for definition 5 is expanded, the expectation of its parts is:

$$\begin{aligned}
\sigma_x^2 &= E(x_i-\mu_x)^2 = E(x_i^2-2x_i\mu_x + \mu_x^2) \\
&= E(x_i^2) - E(2\mu_x x_i) + E(\mu_x^2) \text{ from (4)} \\
&= E(x_i^2) - 2\mu_x E(x_i) + \mu_x^2 \text{ from (1) and (3)} \\
&= E(x_i^2) - (2\mu_x)(\mu_x) + \mu_x^2 \\
&= E(x_i^2) - \mu_x^2.
\end{aligned}$$

Therefore, $E(x_i^2) = \sigma_x^2 + \mu_x^2$. Note that $E(x_i^2) = \sigma_x^2$ when $\mu_x = 0$. Also, as a rule of thumb for finding the variance for a variable X, $E(x_i^2) = \sigma_x^2$ can be used since μ_x drops out of the variance.

Definition 6: $E[(x_i-\mu_x)(y_i-\mu_y)] = \sigma_{xy}$. By definition, the covariance between variables X and Y, σ_{xy}, is the average of the products of their deviations from their means. Thus, $E(x_iy_i) = \sigma_{xy} + \mu_x\mu_y$ which follows from definition 6. If the equation for definition 6 is expanded, the expectation of its parts is:

$$\sigma_{xy} = E[(x_i-\mu_x)(y_i-\mu_y)]$$

$$= E(x_iy_i - \mu_x y_i - \mu_y x_i + \mu_x\mu_y)$$

$$= E(x_iy_i) - \mu_x E(y_i) - \mu_y E(x_i) + \mu_x\mu_y \text{ from (1) and (3)}$$

$$= E(x_iy_i) - \mu_x\mu_y - \mu_y\mu_x + \mu_x\mu_y$$

$$= E(x_iy_i) - \mu_x\mu_y.$$

Therefore, $E(x_iy_i) = \sigma_{xy} + \mu_x\mu_y$. Note that $E(x_iy_i) = \sigma_{xy}$ when either or both μ_x and $\mu_y = 0$.

A general procedure that works well for applying these definitions to determine the expected values of more complicated sums of squares and products of variables is to use the following steps:

Step 1. Substitute elements of the model into the function.

Step 2. Expand the function in terms of the model.

Step 3. Find the expected value of each term of the function.

Step 4. The expected value of the function will be the sum of the expected values of the individual terms.

Example

Let $P_{ij} = \mu + A_i + E_{ij}$, where P_{ij} is an observation on the $j^{\underline{th}}$ record in the $i^{\underline{th}}$ class, μ is a constant, A_i is a variable with $\mu_A = 0$ and variance σ^2_A, E_{ij} is a variable with $\mu_E = 0$ and variance σ^2_E, and the covariance between any two A's, any two E's or any A and any E is zero.

The expected value of any observation is

$$E(P_{ij}) = E(\mu + A_i + E_{ij}) = E(\mu) + E(A_i) + E(E_{ij})$$
$$= \mu + 0 + 0 = \mu.$$

The expected value of any observation squared is:

$$E(P_{ij}^2) = E[(\mu + A_i + E_{ij})^2] = E(\mu^2 + A_i^2 + E_{ij}^2 + 2\mu A_i + 2\mu E_{ij} + 2A_iE_{ij})$$
$$= E(\mu^2) + E(A_i^2) + E(E_{ij}^2) + E(2\mu A_i) + E(2\mu E_{ij}) + E(2A_iE_{ij})$$
$$= \mu^2 + \sigma_A^2 + \sigma_E^2 + 2\mu E(A_i) + 2\mu E(E_{ij}) + 2E(A_iE_{ij})$$
$$= \mu^2 + \sigma_A^2 + \sigma_E^2$$

since $E(A_i)$ and $E(E_{ij})$ both equal zero and $E(A_iE_{ij}) = \sigma_{AE} = 0$.

The expected value of the product of observations ($j' \neq j$) in the same class (class i) is:

$$E(P_{ij}P_{ij'}) = E[(\mu + A_i + E_{ij})(\mu + A_i + E_{ij'})]$$
$$= E(\mu^2 + \mu A_i + \mu E_{ij'} + \mu A_i + A_i^2 + A_iE_{ij'} + \mu E_{ij} + A_iE_{ij} + E_{ij}E_{ij'})$$
$$= \mu^2 + 0 + 0 + 0 + \sigma_A^2 + 0 + 0 + 0 + 0$$
$$= \mu^2 + \sigma_A^2$$

because both $E(A_iE_{ij'})$ and $E(A_iE_{ij})$ are equal to zero when $\sigma_{AE} = 0$ and $E(E_{ij}E_{ij'}) = 0$ when $\sigma_{E_{ij}E_{ij'}} = 0$.

The expected value of the product of observations in different classes (classes i and i') is:

$$E(P_{ij}P_{i'j'}) = E[(\mu + A_i + E_{ij})(\mu + A_{i'} + E_{i'j'})] \quad (i' \neq i' \text{ and } j \neq j' \text{ or } j = j')$$
$$= E(\mu^2 + \mu A_{i'} + \mu E_{i'j'} + \mu A_i + A_iA_{i'} + A_iE_{i'j'} + \mu E_{ij} + A_{i'}E_{ij} + E_{ij}E_{i'j'})$$
$$= \mu^2 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0$$
$$= \mu^2$$

with most terms equal to zero.

Another Example

Suppose a phenotypic observation on animal i is made up of a constant μ, a genetic value G_i, and an environmental effect E_i:

$$P_i = \mu + G_i + E_i$$

where $\mu_G = \mu_E = 0$, $E(G_i^2) = \sigma_G^2$, $E(E_i^2) = \sigma_E^2$, and no covariance between any G's, any E's, and any G with any E.

$$\mu_P = E(P_i) = E(\mu + G_i + E_i) = E(\mu) + E(G_i) + E(E_i)$$
$$= \mu + \mu_G + \mu_E = \mu$$

Note that $\mu_P = \mu$, i.e., two symbols that are equal which will be convenient in the expectation for σ_P^2.

$$E(P_i^2) = E[(\mu + G_i + E_i)^2] = E(\mu^2 + G_i^2 + E_i^2 + 2\mu G_i + 2\mu E_i + 2G_iE_i)$$
$$= \mu^2 + E(G_i^2) + E(E_i^2) + 2\mu E(G_i) + 2\mu E(E_i) + 2E(G_iE_i)$$
$$= \mu^2 + \sigma_G^2 + \sigma_E^2 + 0 + 0 + 0$$

If $\sigma_{GE} \neq 0$, then $E(G_iE_i)$ would also be different from zero.

With no G with E covariance:

$$\sigma_P^2 = E[(P_i-\mu)^2] = E[(\mu + G_i + E_i - \mu)^2] = E[(G_i + E_i)^2]$$
$$= E(G_i^2) + E(E_i^2) + 2E(G_iE_i)$$
$$= \sigma_G^2 + \sigma_E^2 + 0$$

$$E(P_iP_j) = E[(\mu + G_i + E_i)(\mu + G_j + E_j)]$$
$$= E(\mu^2 + \mu G_j + \mu E_j + \mu G_i + G_iG_j + G_iE_j + \mu E_i + G_jE_i + E_iE_j)$$
$$= \mu^2$$

$$COV(P_i,P_j) = E[(P_i-\mu)(P_j-\mu)] = E[(\mu + G_i + E_i - \mu)(\mu + G_j + E_j - \mu)]$$
$$= E[(G_i + E_i)(G_j + E_j)] = E(G_iG_j + G_iE_j + G_jE_i + G_jE_j)$$
$$= 0 + 0 + 0 + 0.$$

TABLE 1.1. SUMMARY OF MEAN AND VARIANCE

	Mean	Variance	Standard Deviation
Symbols:			
Population	μ_x	σ_x^2, V(x)	σ_x
Sample	$\hat{\mu}_x$, $\bar{x}$	$\hat{\sigma}_x^2$, s_x^2	$\hat{\sigma}_x$, s_x
Units	units	units squared	units
Computing formulas:			Alternate Computing Forms for Variance
Nonfrequency data:			
Population	$\frac{\Sigma x_i}{N} = \mu_x$	$\frac{\Sigma(x_i-\mu_x)^2}{N}$	$\frac{\Sigma x_i^2-N\mu_x^2}{N}$
Sample	$\frac{\Sigma x_i}{n} = \hat{\mu}_x = \bar{x}$	$\frac{\Sigma(x_i-\hat{\mu}_x)^2}{n-1}$	$\frac{\Sigma x_i^2-n\hat{\mu}_x^2}{n-1}$; $\frac{\Sigma x_i^2-\frac{(\Sigma x_i)^2}{n}}{n-1}$
Frequency data:			
Population	$\frac{\Sigma f_i y_i}{\Sigma f_i} = \mu_y$	$\frac{\Sigma f_i(y_i-\mu_y)^2}{\Sigma f_i}$	$\frac{\Sigma f_i y_i^2-(\Sigma f_i)\mu_y^2}{\Sigma f_i}$; if $\Sigma f_i = 1$, $\Sigma f_i y_i^2-\mu_y^2$
Sample	$\frac{\Sigma f_i y_i}{\Sigma f_i} = \hat{\mu}_y = \bar{y}$	$\left(\frac{\Sigma f_i (y_i-\hat{\mu}_y)^2}{\Sigma f_i}\right)\left(\frac{n}{n-1}\right)$	As above, but multiply by $\frac{n}{n-1}$

CHAPTER 2

A LITTLE ABOUT MATRIX ALGEBRA

The algebraic description of selection index procedures is very easy with algebra of matrices and vectors. Matrix algebra is also very efficient for writing least squares and mixed model equations and describing the properties of mixed model procedures which will be introduced in the second section of the book. Computer packages are readily available for doing computations interactively using the notation of matrix algebra. Two obvious ones are MATLAB, a personal computer package, which is excellent for working problem sets on a scale not possible with desk calculators, and the IML routines in SAS, a statistical package used in many statistical methods courses. MATLAB is also available on many large computer systems. Although much of the detail of selection index and mixed model procedures is not efficiently done with matrix routines, the arithmetically difficult parts can be illustrated quickly with friendly matrix packages. This chapter will describe the basic rules for matrix algebra and later chapters will describe the computations both element by element and also in matrix form suitable for calculation with a matrix package. MATLAB terminology will be used although IML statements are similar. This chapter and summaries at the end of some chapters on matrix notation can be skipped without missing any of the basic ideas of selection index theory. Many illustrative problems, however, will become more valuable by combining the detail needed for deriving the expected variances and covariances with the ease of doing the final computations with a hands-on matrix package.

The set of numbers such as the coefficients used later in the numerical example of the one-way fixed classification model:

$$C = \begin{pmatrix} 6 & 3 & 1 & 2 \\ 3 & 3 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 2 & 0 & 0 & 2 \end{pmatrix}$$

is called a matrix of 4 rows and 4 columns. Matrices do not have to be square or symmetrical as is C. A symmetrical matrix has columns equal to its corresponding rows. A matrix with only one column is called a vector; e.g., the right-hand sides for that example can be written as the vector, r:

$$r = \begin{pmatrix} y_{..} \\ y_{1.} \\ y_{2.} \\ y_{3.} \end{pmatrix}$$

Notice that elements of matrices and vectors can be represented with numbers or symbols. Matrix arithmetic would require numbers. Matrix algebra may be done with a mixture of both. Matrix algebra is useful in working with and solving least squares and mixed model equations as well as selection index procedures. The notation of matrix algebra is especially convenient and concise for writing simultaneous equations both symbolically and numerically.

The rules of matrix algebra are similar to those for scalar algebra with some important exceptions. Only four rules will be needed for most of this book. Other rules will be introduced when needed.

MATRIX MULTIPLICATION

Rule 1) Matrix multiplication is accomplished by summing the products of each element of each row of the first matrix with the corresponding element of each column of the second matrix (thus the number of elements in each row of the first matrix must equal the number of elements in each column of the second matrix to be **conformable for multiplication**). A new matrix is formed from the sums of these row by column products;

Sum of products of elements of 1st row x 1st column = new element 1,1
Sum of products of elements of 1st row x 2nd column = new element 1,2
.
.
.
Sum of products of elements of 2nd row x 1st column = new element 2,1
Sum of products of elements of 2nd row x 2nd column = new element 2,2
.
.

(The first subscript refers to the row; the second, to the column of the resulting matrix, or vector.)

For example, examine multiplication of a matrix of numbers by column vector of symbols:

$$\begin{pmatrix} 6 & 3 & 1 & 2 \\ 3 & 3 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 2 & 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{A}_1 \\ \hat{A}_2 \\ \hat{A}_3 \end{pmatrix}$$

Sum for 1st row by 1st column;

$$6\hat{\mu} + 3\hat{A}_1 + 1\hat{A}_2 + 2\hat{A}_3$$ (element 1,1) ——> ——> ——> ——>

Sum for 2nd row by 1st column:

$$3\hat{\mu} + 3\hat{A}_1 + 0\hat{A}_2 + 0\hat{A}_3$$ (element 2,1)

——> ——> ——> ——>

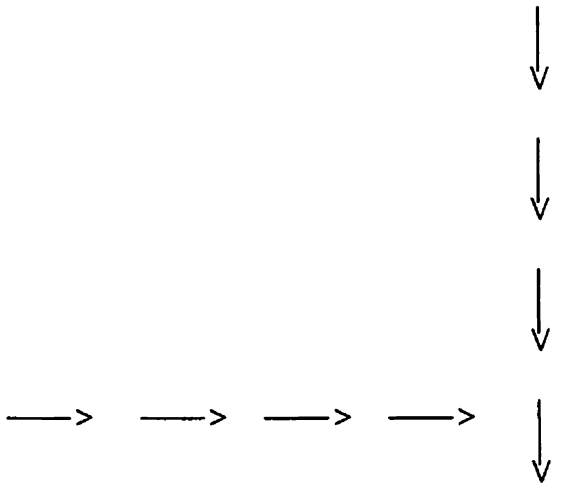
Sum for 3rd row by 1st column;

$$1\hat{\mu} + 0\hat{A}_1 + 1\hat{A}_2 + 0\hat{A}_3$$ (element 3,1)

——> ——> ——> ——>

Sum for 4th row by 1st column;

$$2\hat{\mu} + 0\hat{A}_1 + 0\hat{A}_2 + 2\hat{A}_3 \quad \text{(element 4,1)}$$



The results are the left-hand sides (LHS) of the least squares equations for an example of the one-way classification model. This example is partially numerical, the elements of C, and partially symbolic, the elements of the solution vector:

$$s = \begin{pmatrix} \hat{\mu} \\ \hat{A}_1 \\ \hat{A}_2 \\ \hat{A}_3 \end{pmatrix}$$

The coefficients of the effects on the left-hand sides (LHS) of least squares (LSE) or mixed model equations (MME) make up the coefficient matrix, for example, the matrix C. A similar matrix of coefficients for selection index equations will be denoted P.

The sums on the right of the equal signs make up the right-hand side (RHS) vector:

$$r = \begin{pmatrix} y.. \\ y_{1.} \\ y_{2.} \\ y_{3.} \end{pmatrix}$$

With selection index equations, the RHS vector will be made up of covariances of records in the index with the variable being predicted and may be denoted as g.

Thus, in matrix notation, the set of least square equations can be written:

$$Cs = r \text{ and}$$

the set of selection index equations as:

$$Pb = g$$

where b is the vector of selection index weights.

Multiplication of matrices is a simple extension of matrix by vector multiplication and can be thought of as multiplying the first matrix by a succession of vectors that make up the second matrix. MATLAB would produce the matrix product of A and B and put it in matrix E from E = A*B. Note that usually A*B ≠ B*A. In fact, even if A*B is conformable, B*A might not be conformable for multiplication.

SOLUTIONS WITH INVERSES

Rule 2) If C or P is square and composed of independent rows (columns), the matrix equivalent of division in scalar arithmetic can be used to solve for the solution vector s or b.

In scalar (usual) arithmetic,

$$2x = 4$$

can be solved by premultiplying both sides by the scalar inverse of 2, that is by $(2)^{-1}$;

$$(2)^{-1}(2)x = (2^{-1})(4) = 2 .$$

Because $(2)^{-1}(2) = 1$, then x = 2 is the solution.

In matrix notation, premultiplying both sides by the matrix inverse of C produces the solution vector;

$$C^{-1}Cs = C^{-1}r .$$

If C can be inverted, i.e., is a nonsingular matrix, then;

$$C^{-1}C = I \ .$$

I is the matrix equivalent of the scalar 1. Note that as in scalar algebra, the identity (one) vanishes in multiplication: IC = C, Is = s, etc. In fact, I is a matrix with 1's as diagonal (top left to bottom right) elements and 0's as off-diagonal elements; e.g.,

$$C^{-1}C = I = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Thus, to solve Cs = r, then $C^{-1}Cs = C^{-1}r$ is equivalent to $Is = C^{-1}r$ so that $s = C^{-1}r$.

Note that $C^{-1} \neq I/C$ which has no meaning in matrix algebra.

Finding the elements of the inverse, C^{-1}, from C is usually accomplished by computer programs although students in matrix algebra courses often are required to practice on matrices of order 2×2, 3×3, 4×4, etc. In fact, the command in MATLAB, INV(C) will produce the elements of the inverse.

Note that constraints often must be applied to LSE or MME to make the rows of the coefficient matrix independent so that an inverse of C can be obtained. If the rows are dependent, an inverse does not exist and the matrix is said to be **singular.** A special kind of inverse called the generalized or Penrose inverse can be used in those cases, although care must be taken in interpreting the resulting solutions. In MATLAB, use PINV(C). Thus, in MATLAB, b = INV(P)*g and s = PINV(C)*r.

ADDITION (SUBTRACTION)

Rule 3) Addition (subtraction) of two matrices is accomplished by adding (subtracting) corresponding elements of the two matrices. Thus the matrices must have the same number of rows and columns to be **conformable for addition.** In MATLAB, D = A+B.

SCALAR BY MATRIX MULTIPLICATION

Rule 4) Multiplication of a scalar by a matrix is defined as the multiplication of each element of the matrix by the scalar. If the scalar is $(1-h^2)/h^2$, which for $h^2 = .25$ is (1-.25)/(.25) = 3, then, for example:

$$3I = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

This operation in MATLAB would be:

3*EYE(4) .

CHAPTER 3

QUANTIFYING THE SIMPLE MENDELIAN MODEL

This chapter is not necessary for development of selection procedures for traits influenced by many genes and could be skipped. Nevertheless, examination of simpler models may provide insight into more complicated models which are based on the same principles.

The usual genetic model is

Phenotype = Genotypic effects + Environmental effects

$$P = G + E.$$

The simplest Mendelian model has E = 0 and only three possible genotypes and genotypic effects for one locus with two alleles, A and a.

In a random mating population, if the gene frequency of an allele, A, at a particular locus is p and if there is only one other allele, a, with frequency, 1 - p = q, then the expected frequencies of the three possible genotypes are by the Hardy-Weinberg law

Genotype	Frequency = f_i	Value = y_i
AA	p^2	u
Aa	2pq	[(u+v)/2] + d
aa	q^2	v

Abitrary symbolic effects can be assigned to the genotypes as shown. The value d represents the dominance deviation as a difference of the effect of the heterozygote from the average effect of the homozygotes. There are several possible kinds of dominance depending on the size of d: if d = 0, there is no dominance or equivalently there is lack of dominance or the model contains only additive effects; if d = (u-v)/2, there is complete dominance, that is, the value of Aa equals the value of AA; and if $|d| > (u-v)/2$, there is overdominance, that is, the value of the heterozygote is greater than the value of the AA homozygote or less than the value of the aa homozygote.

POPULATION MEAN

The definition of the population mean or average, μ, is as shown in Chapter 1:

$$\mu = \left(\sum_{i=1}^{n} f_i y_i \right) / \sum_{i=1}^{n} f_i ,$$

where n is the number of different genotypes. Usually, $\Sigma f_i = 1$ but will not if certain genotypes are discarded due to selection.

Application of this formula provides the symbolic mean for the simplest Mendelian model in the case of no selection:

$\mu = v + p(u-v)$, if $d = 0$; and

$\mu = v + p(u-v) + 2pqd$, if $d \neq 0$.

The population average will be maximum when $p = 1$, if $u > v$ and $d \leq (u-v)/2$.

If $d > (u-v)/2$ (overdominance), then the population average will be maximum when $p = \{[(u-v)/2] + d\}/2d$ as can be found by equating the derivative of μ with respect to p to zero.

POPULATION VARIANCE

The definition of the population variance, σ^2, as shown in Chapter 1 is:

$$\sigma^2 = [\sum_{i=1}^{n} f_i(y_i-\mu)^2]/ \sum_{i-1}^{n} f_i$$

If $\Sigma f_i = 1$, then

$$\sigma^2 = \Sigma f_i(y_i-\mu)^2 = \Sigma f_i y_i^2 - \mu^2.$$

Application of this formula will yield equations involving p, u, v, and d, as shown in standard text books on population genetics.

BREEDING VALUE UNDER THE SIMPLE MENDELIAN MODEL

Selection index procedures are primarily aimed toward predicting breeding value. Breeding value can be thought of as the part of the genotypic effects of an animal that can be passed to its progeny. In fact, breeding value is defined for quantitative traits as twice the superiority of an animal that is exhibited in its progeny. The same concept can be shown for effects at a single locus with two alleles.

The frequencies of progeny of the three parental genotypes under random mating are described in the following table.

Parent Genotype	Parent Frequency	Progeny Frequency AA	Aa	aa
AA	p^2	p	q	0
Aa	2pq	p/2	1/2	q/2
aa	q^2	0	p	q

The progeny frequencies are from randomly mating a particular parent type to the rest of the population. For example, Aa × population gives from the gametic arrays the progeny frequencies,

$$[(1/2)(A) + (1/2)(a)] \times [(p)(A) + (q)(a)]$$

$$\downarrow$$

$$(p/2)(AA) + (1/2)(Aa) + (q/2)(aa)\ .$$

The progeny means in symbolic terms for the three parental genotypes are found by applying the formula for the mean with frequencies to be:

$$\mu_{AA} = pu + q\{[(u+v)/2] + d\},$$

$$\mu_{Aa} = .5pu + .5\{[(u+v)/2] + d\} + .5qv,$$

$$\mu_{aa} = p\{[(u+v)/2] + d\} + qv.$$

These expressions show:

(1) that $\mu_{Aa} = (\mu_{AA} + \mu_{aa})/2$ for any values of p and d, that is, the mean for progeny of heterozygotes is the average of the means of the progeny of the two kinds of homozygotes and (2) that breeding values (progeny means) of the parental genotypes depend on gene frequency (even if u>v, μ_{AA} may be less than μ_{aa} when p is small).

HERITABILITY FOR THE SIMPLE GENETIC MODEL

For the simple genetic model (1 locus, 2 alleles, the heterozygote value equal to the average value of the homozygotes, i.e., d = 0) with no environmental effects, the regression of offspring mean on parental value is 1/2 (see formulas in Chapter 1).

Parent	Value	Frequency	Progeny Mean
AA	u	p^2	$pu + q(u+v)/2$
Aa	$(u+v)/2$	$2pq$	$(pu/2) + [(u+v)/4] + (qv/2)$
aa	v	q^2	$p(u+v)/2 + qv$

Here, the mean, $\mu = v + p(u-v)$ and the variance of parents is $pq(u-v)^2/2 = \sigma_g^2$, which is the genetic variance since there are no environmental effects. The covariance of progeny means and parent values is $pq(u-v)^2/4 = \sigma_g^2/2$ and thus the regression of offspring mean on parent value is 1/2.

The following principles should be noted:

1. Any kind of dominance will decrease the regression coefficient.
2. Selection on parents will not affect the regression if the heterozygote has a value which is the average of the values of the homozygotes (additive model).
3. If there is some form of dominance, selection on the parents will, in general, affect the regression as can be seen by plotting progeny means against parental values.

SIMPLE GENETIC MODEL WITH ENVIRONMENTAL EFFECTS

Suppose to the simple genetic model that a random environmental contribution is added that averages zero but has variance σ_e^2. Then, phenotype = genotype + environment, or $P = G + E$. If there is no correlation between G and E, then phenotypic variance, $\sigma_p^2 = \sigma_g^2 + \sigma_e^2$. In terms of selection, usually the components of P cannot be separated directly. The environmental effect, E, may mask what is to be evaluated, G.

HERITABILITY DEFINED

Heritability is defined in the "broad sense" as the ratio of all of the genetic variance to the total variance;

$$h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_e^2).$$

With additive gene action (the heterozygote intermediate in value between homozygotes) for the simple model the covariance between parent and progeny is $(1/2)(\sigma^2_{g_A})$. This result can be shown to be true even if environmental variation exits because the environmental variation is assumed to be random with average value of zero.

If d does not equal 0 (some form of dominance), part of the genetic variance will be due to the dominance effects, $\sigma^2_{g_D}$ and some to additive effects, $\sigma^2_{g_A}$.

If there is some form of dominance, the regression of progeny on parent is reduced. The covariance, however, between progeny and parent is $(1/2)(\sigma^2_{g_A})$ either with dominance or with no dominance (see appendix this chapter).

Heritability is defined in the "narrow sense" as the ratio of additive genetic variance to the total variance;

$$h^2 = \frac{\sigma^2_{g_A}}{\sigma^2_g + \sigma^2_e}, \qquad \text{where} \quad \sigma^2_g = \sigma^2_{g_A} + \sigma^2_{g_D}.$$

Thus, twice the regression of progeny mean on parent value equals heritability in the "narrow sense" even with dominance in the simple genetic model with random environmental effects. The same will be true for quantitative traits influenced by genes at many loci.

Later chapters show that additive genetic effects are most important since they have a much greater chance than dominance or epistatic effects of being transmitted from one generation to the next.

APPENDIX TO CHAPTER THREE

DERIVATION OF σ^2_g, $\sigma^2_{g_A}$, $\sigma^2_{g_D}$ FOR 1 LOCUS WITH 2 ALLELES

Let value of AA = 1, value of Aa = d, and value of aa = 0. The frequency of A is p and the frequency of **a** is 1 - p = q. Random mating is assumed.

Total Genetic Variance, σ^2_g

$$\mu_g = p^2 + 2pqd$$

$$\sigma^2_g = p^2 + 2pqd^2 + 0 - (p^2 + 2pqd)^2$$

$$= p^2 + 2pqd^2 - p^4 - 4p^3qd - 4p^2(q^2)d^2$$

$$= pq[p(1+p) + 2d(d - 2p^2 - 2pqd)]$$

If $d = 1/2$, $\sigma^2_g = pq/2$.

Regression of Genotypic Value, G, on Number of "+" Genes, X, to Define $\sigma^2_{g_A}$ (Depends on p)

The following table describes frequencies and genetic values associated with number of positive genes in the genotype:

Genotype	Frequency	G	X
AA	p^2	1	2
Aa	2pq	d	1
aa	q^2	0	0

$\sigma^2_{g_A} = r^2_{gx}\sigma^2_g$, the additive genetic variance, is defined as the variance in G due to additive gene effects (i.e., variance in G accounted for by regression of G on X):

$$\sigma^2_{g_A} = \frac{\sigma_{gx}\sigma_{gx}}{\sigma^2_x}$$

Now evaluate σ_{gx} and σ_x^2.

$$\mu_x = 2p$$

$$\sigma_x^2 = p^2(2)^2 + 2pq + 0 - (2p)^2$$

$$= 4p^2 + 2pq - 4p^2 = 2pq$$

$$\mu_g = p^2 + 2pqd$$

$$\sigma_{gx} = 2p^2 + 2pqd + 0 - 2p(p^2 + 2pqd)$$

$$= 2p^2 + 2pqd - 2p^3 - 4p^2qd$$

$$= 2pq(p+d-2pd)$$

Thus

$$\sigma_{gA}^2 = \frac{(\sigma_{gx})^2}{\sigma_x^2} = \frac{[2pq(p+d-2pd)]^2}{2pq}$$

$$= 2pq(p+d-2pd)^2$$

$$= 2pq(p^2 + d^2 + 4p^2d^2 + 2pd - 4p^2d - 4pd^2).$$

The dominance genetic variance is variance in G not accounted for by regression on X:

$$\sigma_{gD}^2 = \sigma_g^2 - \sigma_{gA}^2$$

$$= p^2q[1-p+4d(d-pd-1+p)].$$

If $d = 1/2$, $\sigma_{gA}^2 = \sigma_{gD}^2 = 0$. With this assignment of genetic values, $d = 1/2$ corresponds to no dominance, i.e., $(v_{AA} + v_{aa})/2 = (1+0)/2 = 1/2$.

COVARIANCE (PROGENY, PARENT) WITH DOMINANCE

The table summarizes parental frequencies and parent and progeny values when the genotypic values are defined as 1, d, and 0.

		Values	
Parents	Frequency	Parent	Progeny Mean
AA	p^2	1	p + qd
Aa	2pq	d	(1/2)(p) + (1/2)(d)
aa	q^2	0	pd

$$\mu_{progeny} = \mu_{parent} = \mu_{population} = p^2 + 2pqd$$

$$Cov = p^2(p + qd) + 2pqd(1/2)(p+d) - (p^2 + 2pqd)^2$$

$$= pq(p + d - 2pd)^2,$$

which is $(1/2)(\sigma^2_{g_A})$ no matter what the values of p and d are.

In this derivation, the values of the genotypes (1, d, and 0) have been scaled from general phenotypic values of u, [(u+v)/2] + d', and v by subtracting v from each general value and then dividing by u-v. Note that the scaled d = (1/2) + [d'/(u-v)] in terms of the general values. To convert the above results (variances) back to general values, multiply by $(u-v)^2$.

CHAPTER 4

A SHORT SUMMATION ON POPULATION GENETICS

The principles of quantitative genetics and population genetics are closely linked. Population genetics is primarily concerned with identifiable alleles and their frequencies whereas quantitative genetics is primarily concerned with small effects of many unidentifiable alleles. Causes and effects of changes in frequencies of alleles is a major common concern of these two related fields of genetics. This chapter summarizes a few of the more obvious results from population genetics. Texts such as Doolittle (1987), Hartl (1980), Falconer (1989), provide much more complete development of the principles of population genetics. This chapter could be expanded in lectures if the topics of quantitative and population genetics are covered in one course. This chapter should be skipped if the course is limited to selection for quantitative traits.

THE HARDY-WEINBERG LAW

If in a large population, **p** is the frequency of gene **A** and **q** is the frequency of the other allele, **a**, then after one generation of random mating the genotypes will have and will continue to have in future generations the frequencies $\mathbf{p}^2$ for **AA**, **2pq** for **Aa**, and $\mathbf{q}^2$ for **aa**. Note that $\mathbf{p} + \mathbf{q} = 1$. Hence, $\mathbf{q} = 1 - \mathbf{p}$.

This principle can be extended to the case of n alleles, A_i, (i=1, ..., n), with frequencies p_i, by computing the frequencies of the genotypes obtained from multiplying the gametic array for males by the same gametic array for females, $(p_1A_1 + ... + p_nA_n)$. The genotypes and their frequencies will be:

homozygotes: A_iA_i with frequencies p_i^2 for i = 1, ..., n;

heterozygotes: A_iA_j with frequencies $2p_ip_j$ for all $i \neq j$.

Again, $\sum_{i=1}^{n} p_i = 1$.

ESTIMATION OF GENE FREQUENCIES

The general formula for the

$$\text{frequency of some allele} = \frac{\text{Number of that allele}}{\text{Total number of genes at that locus}}.$$

The problems of estimation are illustrated in the following special cases.

(1) Dominance:

The frequency of a recessive gene in a random mating population can be estimated from the knowledge that a fraction q^2 of the population is expected to be homozygous recessive. Then,

$$q = \sqrt{\text{number recessive types/total number of animals}},$$

and $p = 1 - q$.

In the case of multiple alleles with complete dominance, the frequency of the most recessive allele is estimated first. For example, suppose A_1 is dominant to A_2 and A_3, and A_2 is dominant to A_3 as shown below.

Genotype	Phenotype	Expected Frequency
A_1A_1 A_1A_2 A_1A_3	A_1	$p_1^2 + 2p_1p_2 + 2p_1p_3$
A_2A_2 A_2A_3	A_2	$p_2^2 + 2p_2p_3$
A_3A_3	A_3	p_3^2

Then from the last equation, $\mathbf{p}_3 = \sqrt{\text{number } A_3 \text{ type/total number}}$. Put the estimate of $\mathbf{p}_3$ into the second equation, $\mathbf{p}_2^2 + 2\mathbf{p}_2\mathbf{p}_3$ = number $\mathbf{A}_2$ type/total number, and solve for $\mathbf{p}_2$. Next substitute the estimates of $\mathbf{p}_3$ and $\mathbf{p}_2$ into the first equation, $\mathbf{p}_1^2 + 2\mathbf{p}_1\mathbf{p}_2 + 2\mathbf{p}_1\mathbf{p}_3$ = number $\mathbf{A}_1$ type/total number and solve for $\mathbf{p}_1$ or find $\mathbf{p}_1$ by difference since $\mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_3 = 1$, or $\mathbf{p}_1 = 1 - \mathbf{p}_2 - \mathbf{p}_3$.

(2) Incomplete dominance:

With incomplete dominance, heterozygotes can be distinguished from homozygotes so that the gene frequencies can be found from the general formula whether or not the population is randomly mating. For example, with three alleles,

$$\mathbf{p}_1 = \frac{\text{Number } A_1 \text{ alleles}}{\text{Total number of genes at the A locus}}$$

Each $\mathbf{A}_1\mathbf{A}_1$ genotype contributes two $\mathbf{A}_1$ alleles; each $\mathbf{A}_1\mathbf{A}_2$ genotype contributes one $\mathbf{A}_1$ allele; and each $\mathbf{A}_1\mathbf{A}_3$ genotype contributes one $\mathbf{A}_1$ allele to the number of $\mathbf{A}_1$ genes. Then,

$$\mathbf{p}_1 = \frac{2(\text{number of } A_1A_1) + \text{number of } A_1A_2 + \text{number of } A_1A_3}{2(\text{total number of animals})},$$

and $\mathbf{p}_2$ and $\mathbf{p}_3$ may be estimated similarly.

FREQUENCIES OF COMPOSITE GENOTYPES

The frequencies of composite genotypes in a random mating population which is at equilibrium with respect to linkage are equal to the products of the single locus frequencies. For example, if the frequencies of $\mathbf{A}_1$ and $\mathbf{A}_2$ alleles are $\mathbf{p}_1$ and $\mathbf{p}_2$, respectively, and the frequencies of $\mathbf{B}_1$ and $\mathbf{B}_2$ alleles are $\mathbf{r}_1$ and $\mathbf{r}_2$, respectively, then the frequencies of the composite (two-locus) genotypes at equilibrium will be:

Genotypes	Frequencies
$A_1A_1B_1B_1$	$p_1^2r_1^2$
$A_1A_1B_1B_2$	$p_1^2(2r_1r_2)$
$A_1A_1B_2B_2$	$p_1^2r_2^2$
$A_1A_2B_1B_1$	$(2p_1p_2)r_1^2$
$A_1A_2B_1B_2$	$(2p_1p_2)(2r_1r_2)$
$A_1A_2B_2B_2$	$(2p_1p_2)r_2^2$
$A_2A_2B_1B_1$	$p_2^2r_1^2$
$A_2A_2B_1B_2$	$p_2^2(2r_1r_2)$
$A_2A_2B_2B_2$	$p_2^2r_2^2$

Extension to more than two alleles per locus or more loci follows the same pattern.

EFFECT OF SELECTION ON GENE FREQUENCIES

Selection may change the frequency of a certain allele in a population. Gene frequency after selection (among the survivors and with random mating of the survivors among the next generation) depends on the fitness of the genotypes and allelic frequencies in the current generation. Fitness of a genotype is defined as the proportion of the genotype

that reproduces relative to the other genotypes. Let **s** be the fraction of AA genotypes, **r** be the fraction of Aa genotypes, and **t** the fraction of **aa** genotypes that do not reproduce, where $1 \geq s, r$ and $t \geq 0$. By counting alleles, the frequency of allele **a** after selection is expected to be

$$q_1 = \frac{\text{number of "a" genes among survivors}}{2(\text{number of survivors})}.$$

This expression has specific forms for special kinds of selection as described later. The change in allelic frequency from one generation to another is the difference in allelic frequencies between the generations, i.e.,

$$\Delta q = q_n - q_{n-1},$$

where the subscripts refer to generations **n** and **n**-1.

Special Cases

In the following special cases, some simplifications may be made.

(1) No homozygous recessive individuals reproduce (zero fitness for the aa genotype; $s = 0, r = 0, t = 1$).

The composition of the initial generation ($n = 0$) can be described:

Genotype	Frequency	Fitness	Relative Frequency of Survivors
AA	p^2	1-s = 1	p^2
Aa	2pq	1-r = 1	2pq
aa	q^2	1-t = 0	0
Total	1		$p^2 + 2pq = 1 - q^2$

Then, by the general equation because only the **Aa** genotype contains an **a** allele

$$q_1 = \frac{\text{number of a alleles in survivors}}{\text{total number alleles in survivors}},$$

$$q_1 = \frac{2pq(\text{number of animals})}{2(p^2 + 2pq)(\text{number of animals})} = \frac{q}{1+q}$$

If this procedure is followed through n generations, the frequency of the allele **a** will be $q_n = q/(1 + nq)$, where **q** was the original allelic frequency of the recessive allele. From this expression, the number of generations, **n**, required to go from an allelic frequency of **q** to one of q_t is found to be $n = (1/q_t) - (1/q)$.

(2) Selection in favor of heterozygotes ($r = 0$)

The composition of the initial generation ($n = 0$) before and after selection is:

Genotype	Frequency	Fitness	Relative Frequency of Survivors
AA	p^2	1-s	$p^2(1-s)$
Aa	2pq	1	2pq
aa	q^2	1-t	$q^2(1-t)$
Total	1		$1-sp^2-tq^2$

Application of the general procedure for finding the new allelic frequency, q_1, gives

$$q_1 = \frac{(1-t)q^2 + pq}{1-sp^2-tq^2}.$$

The change in allelic frequency from the zero generation to the next is:

$$\Delta q = \frac{pq(sp-tq)}{1-sp^2-tq^2}.$$

When $\Delta q = 0$, there will be no change in allelic frequency from the (n-1)st generation to the n^{th} generation and the population will be at equilibrium. As seen from the numerator, equilibrium occurs when $sp - tq = 0$. Thus, equilibrium allelic frequency will be reached when $p = t/(s+t)$ and $q = s/(s+t)$.

(3) Partial selection against homozygous recessives ($s = 0$, $r = 0$, $t > 0$)

The composition of the initial generation before and after selection is:

Genotype	Frequency	Fitness	Relative Frequency of Survivors
AA	p^2	1	p^2
Aa	2pq	1	2pq
aa	q^2	1-t	$q^2(1-t)$
Total	1		$1-tq^2$

The allelic frequency, q_1, in the survivors is by the general procedure

$$q_1 = \frac{q(1-tq)}{1-tq^2}, \text{ and}$$

$$\Delta q = q_1 - q = \frac{-tq^2(1-q)}{1-tq^2}.$$

(4) Selection against heterozygotes (**s** = 0, **r** > 0, **t** = 0)

The composition of the initial generation before and after selection is:

Genotype	Frequency	Fitness	Relative Frequency of Survivors
AA	p^2	1	p^2
Aa	2pq	1-r	2pq(1-r)
aa	q^2	1	q^2
Total	1		1-2pqr

By the general procedure, the allelic frequency, $\mathbf{q}_1$, in the survivors is

$$\mathbf{q}_1 = \frac{q(1-rp)}{1-2rpq}, \text{ and}$$

$$\Delta\mathbf{q} = q_1 - q = \frac{rpq(2q-1)}{1-2rpq}.$$

(5) Changes in allelic frequencies with other combinations of fitness values can be worked out similarly by the general procedure.

DETECTION OF CARRIERS OF RECESSIVE ALLELES

The confidence of detection of a heterozygote for a recessive allele (i.e., a carrier of the recessive gene) depends on the probability of obtaining **at least one** affected offspring in **n** offspring if the suspected carrier is actually a carrier. This probability is one minus the probability of obtaining **all** normal offspring in **n** offspring.

A general testing procedure is to mate a suspected carrier to a group of females which produce a fraction, **p**, **A** alleles and a fraction, **q**, **a** alleles. Then, if the suspect is really a carrier, the probability that all **n** offspring will be normal is $[1 - q/2)]^n$ and the confidence of detection (i.e., proving the animal is a carrier) is $1 - [1 - (q/2)]^n$. Usually testing is for males because of their potentially high reproductive rate relative to females.

Some special cases are:

(1) Mating a carrier male to known homozygous recessive females

The probability of obtaining all normal offspring is $(1/2)^n$.

Therefore, the "confidence" of detecting him as a carrier is $1 - (1/2)^n$.

Note that $\mathbf{q} = 1$ and that $1 - (\mathbf{q}/2) = 1/2$.

(2) Mating a carrier male to known carrier (heterozygous) females

The "confidence" of detecting him as a carrier is $1 - (3/4)^n$.

Note that $\mathbf{q} = 1/2$ and that $1 - (\mathbf{q}/2) = 3/4$.

(3) Mating a carrier male to his own daughters.

The "confidence" of detecting him as a carrier is $1 - (7/8)^n$. Note that $\mathbf{q} = 1/4$ and that $1 - (\mathbf{q}/2) = 7/8$. This probability is calculated under the assumption that the dams were all homozygous for the normal allele. With this system, recessive alleles at all loci will have the same chance of detection.

(4) Mating a carrier male at random in a population where the frequency of the recessive gene is $\mathbf{q}^*$ in the previous generation

The "confidence" of detecting him as a carrier is $1 - [(2+q^*)/2(1+q^*)]^n$.

Note that $\mathbf{q} = \mathbf{q}^*/(1+\mathbf{q}^*)$ since none of the homozygous recessive females will be mated.

What should be remembered about testing for carriers is that even one verified affected offspring marks a suspected carrier as a carrier. Even if all offspring are normal, that will never completely rule out the possibility a male is a carrier, even though the probability of detection may be quite high.

Table 4.1 shows the confidence of detection of carrier males for the four testing systems. A further discussion of method 4 follows.

TABLE 4.1. CHANCES OF DETECTING A CARRIER MALE FOR VARIOUS TYPES OF MATINGS

Number of Progeny n	Detects only one lethal: homozygous recessive females $1-(1/2)^n$	Detects only one lethal: known carrier females $1-(3/4)^n$	Detects all lethals carried: own daughters $1-(7/8)^n$	Detects all lethals depending on frequency random in population $1-\left(\frac{2+q^*}{2(1+q^*)}\right)^n$: Lethal Gene Frequency = q^* in previous generation .2	.1	.05	.01	.001
1	.50	.25	.12	.08	.05	.02	.00	.00
2	.75	.44	.23	.16	.09	.05	.01	.00
3	.88	.58	.33	.23	.13	.07	.01	.00
4	.94	.68	.41	.29	.19	.09	.02	.00
5	.97	.76	.49	.35	.21	.11	.02	.00
6	.98	.82	.55	.41	.24	.13	.03	.00
7	.99	.87	.61	.46	.28	.16	.03	.00
8	1.00	.90	.66	.50	.31	.18	.04	.00
9		.92	.70	.54	.34	.20	.04	.00
10		.94	.74	.58	.37	.21	.05	.00
15		.99	.87	.73	.50	.30	.07	.01
20		1.00	.93	.82	.61	.38	.09	.01
50			1.00	.99	.90	.70	.22	.02
100				1.00	.99	.91	.39	.05
200					1.00	.99	.63	.10
300						1.00	.77	.14
400							.86	.18
500							.92	.22

ARTIFICIAL INSEMINATION AND UNDESIRABLE RECESSIVES

Method 4 of the preceding section can be used to decrease the frequency of all undesirable genes by progeny testing all males at random in the population before heavy use through artificial insemination. The following describes how such a program would work for dairy cattle where AI is widely used.

The essential question is, "Can AI be used to find carrier bulls before they spread undesirable genes?" The answer is yes since any good young sire sampling program will provide for each young bull producing at least 200 progeny. In the dairy situation, 200 calves may yield 50 or so production-tested daughters. At the same time, the 200 progeny will provide an excellent test of whether the bull is a carrier of any undesirable recessive genes.

What does this mean in terms of numbers of affected calves? An AI program which observes 200 tested progeny can be compared with what would happen without AI. The effect of AI testing with 200 progeny versus no testing is shown in the Table 4.2.

With all the expressed fears that AI may sabotage a population by spreading an undesirable allele throughout the population, it is more than a little reassuring to know this is unlikely to happen. More reassuring is the knowledge that a properly set up young sire sampling program in AI will actually protect a population against undesirable genes and reduce the number of affected calves.

A more technical description follows on how to calculate expected frequencies of affected calves of future generations with an AI testing scheme with various numbers of test matings and initial gene frequencies.

TABLE 4.2. NUMBER OF AFFECTED OFFSPRING WITH NO PROGENY TESTING AND WITH TESTING WITH 200 PROGENY

	Before Testing	No. of affected progeny per million progeny born in generation						
		1	2	3	4	5	---	10
No testing	250,000	111,111	62,500	40,000	27,778	20,408	---	6,944
AI test	250,000	0	0	2	28	55	---	9
No testing	40,000	27,778	20,408	15,625	12,346	10,000	---	4,444
AI test	40,000	0	2	28	55	43	---	7
No testing	10,000	8,264	6,944	5,917	5,102	4,444	---	2,500
AI test	10,000	1	23	55	46	29	---	6
No testing	2,500	2,268	2,066	1,890	1,736	1,600	---	1,111
AI testing	2,500	20	54	47	30	20	---	4
No testing	100	98	96	94	92	91	---	83
AI testing	100	37	23	16	11	8	---	3

THE EFFECT OF TESTING BULLS IN AI ON THE FREQUENCY OF RECESSIVE ALLELES

As the frequency of a recessive allele drops under the conditions of AI, the confidence of detecting a carrier by random mating goes down. What will be the effect of the reduced confidence on selection against the gene?

The solution can be obtained by computing the allelic frequencies for several generations. Males will be progeny-tested on n females. All males and females which are homozygous for recessive alleles will be culled. Heterozygotes have the same fitness as the "normal" homozygotes. Let $\mathbf{p_j}$ = frequency of the normal allele, **A**, in males surviving selection, $\mathbf{q_j}$ = frequency of the other allele, **a**, in males surviving selection, $\mathbf{P_j}$ = frequency of **A** in females surviving selection, $\mathbf{Q_j}$ = frequency of **a** in females surviving selection, and **j** is the generation number. The frequency of genotypes in the next generation can be found by expanding $(\mathbf{p_jA} + \mathbf{q_ja})(\mathbf{P_jA} + \mathbf{Q_ja})$. The composition of the next generation before and after selection is:

	Males			Females		
Genotype	Frequency	Fitness	Frequency Survivors	Frequency	Fitness	Frequency Survivors
AA	p_jP_j	1	p_jP_j	p_jP_j	1	p_jP_j
Aa	$p_jQ_j+q_jP_j$	α_i	$\alpha_i(p_jQ_j+q_jP_j)$	$p_jQ_j+q_jP_j$	1	$p_jQ_j+q_jP_j$
aa	q_jQ_j	0	0	q_jQ_j	0	0
Total	1		$p_jP_j+\alpha_i(p_jQ_j+q_jP_j)$	1		$p_j+q_jP_j$

$\alpha_i(i=j+1)$ is the probability of not detecting a carrier by random mating to the population. Males are tested in the population of contemporary females. $\alpha_i = (1 - .5Q_{j+1})^n$, where **n** is the number of progeny and Q_{j+1} is the frequency of **a** among the surviving females. The frequency of a among the selected males is

$$q_{j+1} = (\alpha_i/2)(p_jQ_j+q_jP_j)/[p_jP_j + \alpha_i(p_jQ_j+q_jP_j)] .$$

The composition of the next generation can be found by expanding

$$(p_{j+1}A + q_{j+1}a)(P_{j+1}A + Q_{j+1}a) .$$

Note that this is a repeating pattern and can be easily programmed for a computer.

CHAPTER 5

GENES IDENTICAL BY DESCENT--THE BASIS OF GENETIC LIKENESS

Individuals may have genes in common from a common ancestor. Such genes are identical by descent. If genes are identical but not necessarily from a common ancestor, they are identical in state. The term allele may be more appropriate but the term gene will be used here.

The concept of identity by descent is an approach to the complications of multi-allelic, multi-loci gene systems which affect quantitative traits. With the identity by descent approach, there is no need to know how many alleles are at a locus, the value of each allele, the number of loci which have genes influencing the quantitative trait, or the gene frequencies. This approach was formulated by Malécot (1948) and about the same time by C. C. Cockerham and C. R. Henderson, who further developed the concept. The identity by descent approach is to calculate probabilities of genes, genotypes and non-allelic combinations of genes being identical because of common ancestors.

Two limitations of the probabilistic method are:

1. Calculations of probabilities must begin at a specified base period even though most life probably originated from a small number of genes.
2. The method estimates how many genes are identical by descent between two animals only on a probability basis.

The notation is that an animal will have genes b_ib_j at the **b** locus where the subscript describes the origin of the gene. The basis for calculation of relationships is the probability that a random gene at any locus, say the **b** locus, is identical by descent for a pair of individuals. At some arbitrary base period, the **b** genes of the common ancestor are tagged and then the probability that the **b** genes of the two individuals will be common by descent is computed.

Let the genotypes of two animals at the **b** locus be b_ib_j and b_mb_n where the subscripts refer to the origin of the gene. The probability that the genes at a locus are identical by descent between two individuals is defined by comparing the origins of the first gene of the first animal with the first and second genes of the second animal, and the second gene of the first animal with the first and second genes of the second.

Thus, for the four possible combinations:

$$\text{Probability } (b_i = b_m) = 0 \text{ if } i \neq m; = 1 \text{ if } i = m,$$

$$\text{Probability } (b_i = b_n) = 0 \text{ if } i \neq n; = 1 \text{ if } i = n,$$

$$\text{Probability } (b_j = b_m) = 0 \text{ if } j \neq m; = 1 \text{ if } j = m,$$

$$\text{Probability } (b_j = b_n) = 0 \text{ if } j \neq n; = 1 \text{ if } j = n.$$

The probability that a random gene at this locus is identical in two animals is the average probability for these four comparisons, i.e.,

$$[P(i=m) + P(i=n) + P(j=m) + P(j=n)]/4.$$

In fact, this expression is the same as the probability that a random gene from one animal and a random gene from the other animal will be identical by descent.

As a specific example, suppose that two unrelated noninbred animals are mated, i.e., the mating is $b_1b_2 \times b_3b_4$. The possible offspring are b_1b_3, b_1b_4, b_2b_3, b_2b_4. The fraction

of random pairs of genes being identical between any progeny, say b_1b_3, and any parent, say b_1b_2, is:

$$P(b_1 = b_1) = 1$$

$$P(b_1 = b_3) = 0$$

$$P(b_2 = b_1) = 0$$

$$P(b_2 = b_3) = 0$$

with the average being 1/4.

ADDITIVE RELATIONSHIP

The relationship of an individual with itself is considered generally to be one. The "a" or additive relationship between two individuals is defined as twice the fraction of genes identical by descent so that the additive relationship of a noninbred animal with itself is one. As shown in the appendix to Chapter 6, because each locus has two additive gene effects, the additive relationship is the measure of the fraction of additive gene effects in common between relatives. In a noninbred population, the additive relationship is equal to the coefficient of relationship. The coefficient of relationship is also equal to the correlation between additive effects as will be described in Chapter 6.

The coefficient of relationship between animals i and j is:

$$r_{ij} = a_{ij}/\sqrt{a_{ii}a_{jj}},$$

where a_{ij} is the additive relationship between i and j, a_{ii} is the additive relationship of i to itself ($a_{ii} = 1$ if noninbred) and a_{jj} is the additive relationship of j to itself. Thus, the additive relationship is sometimes called the numerator relationship because the additive relationship is the numerator for the coefficient of relationship. The following table

describes the probabilities for most common kinds of comparisons for pairs of individuals.

PROBABILITIES OF GENES IDENTICAL BY DESCENT

Comparison	Fraction Identical by Descent	Additive Relationship
b_1b_1 with b_1b_1 (completely inbred with self)	1	2
b_1b_1 with b_1b_2	1/2	1
b_1b_1 with b_2b_2	0	0
b_1b_2 with b_1b_2 (noninbred with self)	1/2	1
b_1b_2 with b_1b_3	1/4	1/2
b_1b_2 with b_1b_4	1/4	1/2
b_1b_2 with b_3b_4	0	0

Parent-progeny Relationship

Unrelated and noninbred parents, b_1b_2 and b_3b_4, have potential progeny b_1b_3, b_1b_4, b_2b_3, and b_2b_4. From the table above, the fraction of genes identical by descent for any one parent with b_1b_3 is 1/4; with b_1b_4, 1/4; with b_2b_3, 1/4; and with b_2b_4, 1/4. The average is $(1/4 + 1/4 + 1/4 + 1/4)/4 = 1/4$ and the additive relationship is 1/2.

Grandparent-grandprogeny Relationship

Two unrelated and noninbred animals, b_1b_2 and b_3b_4, have potential progeny b_1b_3, b_1b_4, b_2b_3, and b_2b_4. One of these progeny chosen at random, say b_1b_3, is mated to an unrelated animal, b_5b_6, chosen from the population. The potential genotypes of their

progeny are b_1b_5, b_1b_6, b_3b_5, and b_3b_6. Now compare genes of either grandparent, say b_1b_2, with genes of the grandprogeny. The fraction of genes of that grandparent that are identical by descent with b_1b_5 is 1/4; with b_1b_6, 1/4; with b_3b_5, 0; and with b_3b_6, 0. The average is 1/8.

The same average would be found for the grandparent, that is b_1b_2, with the other 12 possible grandprogeny types. In one-half the comparisons the grandprogeny and grandparent are unrelated in the sense that no genes are alike at that locus. Since the probability of no genes in common at one loci is 1/2, the probability of no genes in common at **n** loci is $(1/2)^n$ for grandparent and grandprogeny pairs which is not a very large probability, even for number of loci as small as four. The average identical by descent over all loci is likely to be quite close to the calculated probability of genes being identical by descent.

Full sib Relationship

Two unrelated and noninbred animals, b_1b_2 and b_3b_4, have progeny b_1b_3, b_1b_4, b_2b_3, and b_2b_4. When the full sib progeny are randomly compared, there are 16 different combinations of pairs of full sibs, each having equal frequency. The values in the table are the probabilities of genes being identical for each of the 16 comparisons.

		Possible Genotypes of 2nd Full Sib with Frequencies			
		1/4 b_1b_3	1/4 b_1b_4	1/4 b_2b_3	1/4 b_2b_4
Possible	1/4 b_1b_3	1/2	1/4	1/4	0
Genotypes of	1/4 b_1b_4	1/4	1/2	0	1/4
1st full sib	1/4 b_2b_3	1/4	0	1/2	1/4
with frequencies	1/4 b_2b_4	0	1/4	1/4	1/2

The average will be $\Sigma f_i X_i$. For all 16 cells, $f_i = 1/16$. The average fraction of genes identical by descent = (1/16)[(4)(1/2) + (8)(1/4) + (4)(0)] = 1/4 as before. Note that although the average fraction of genes identical by descent is 1/4 that 1/4 of the comparisons have probability 1/2 (an identical genotype), 1/2 have probability 1/4, and 1/4 have probability 0. One-fourth of the comparisons have no genes in common at one locus, and therefore the probability of no genes in common at **n** loci = $(1/4)^n$ for full sibs.

Half-sib Relationship

When animal b_1b_2 is mated to b_3b_4, they have potential progeny b_1b_3, b_1b_4, b_2b_3, and b_2b_4 with equal frequencies. When animal b_1b_2 is also mated to b_5b_6, they have potential progeny b_1b_5, b_1b_6, b_2b_5, and b_2b_6. The values in the table are fractions of genes identical by descent for each of the 16 possible pairs of half-sibs.

		Possible Genotypes of 1st Half-Sib with Frequencies			
		1/4 b_1b_3	1/4 b_1b_4	1/4 b_2b_3	1/4 b_2b_4
Possible	1/4 b_1b_5	1/4	1/4	0	0
Genotypes of	1/4 b_1b_6	1/4	1/4	0	0
2nd Half-Sib	1/4 b_2b_5	0	0	1/4	1/4
with frequencies	1/4 b_2b_6	0	0	1/4	1/4

The average fraction of genes identical by descent is:

(1/16)[(8)(1/4) + (8)(0)] = 1/8, and the additive relationship is 1/4. One-half of the comparisons have no genes in common at one locus, and the probability of no genes in common at **n** loci = $(1/2)^n$.

Summary of Most Common Relationships

Relationship	Ave. Fraction Identical	Probability of Genotype Identical	Probability No Genes Identical at **n** Loci
Parent-progeny	1/4	0	0
Grandparent-grandprogeny	1/8	0	$(1/2)^n$
Full sibs	1/4	1/4	$(1/4)^n$
Half sibs	1/8	0	$(1/2)^n$

DOMINANCE RELATIONSHIP

The probability of an identical genotype at one locus by descent is the probability that the pair of genes at one locus for two animals identical by descent, i.e., for relatives with genotypes b_ib_j and b_mb_n, P(genotype identical) = $P(b_ib_j = b_mb_n)$. The only pairs in the above table that can have a genotype at one locus identical by descent are pairs of full sibs, e.g., b_1b_4 with b_1b_4. The dominance relationship between a pair of animals is defined as the probability of genotypes being identical by descent.

The following is an example of computing average probability of genotypes in common for full sibs.

Let the parents be unrelated so that their symbolic genotypes and those of their full sib progeny can be represented as:

Parents: b_1b_2, b_3b_4

Full Sib Progeny (with frequencies):

$1/4\ b_1b_3$, $1/4\ b_1b_4$, $1/4\ b_2b_3$, $1/4\ b_2b_4$

The average probability that $b_ib_j = b_mb_n$ is the average of all 16 comparisons as shown in the following table.

		Possible Genotypes of 2nd Full Sib with Frequencies			
		1/4 b_1b_3	1/4 b_1b_4	1/4 b_2b_3	1/4b_2b_4
Possible Genotypes of 1st Full Sib with Frequencies	1/4 b_1b_3	The frequency of each comparison is (1/4)(1/4) = 1/16.			
	1/4 b_1b_4				
	1/4 b_2b_3				
	1/4 b_2b_4				

Then, the average P(genotypes identical at the "**b**" locus) is:

$$= (1/16)[P(b_1b_3=b_1b_3) + P(b_1b_3=b_1b_4) + P(b_1b_3=b_2b_3) + P(b_1b_3=b_2b_4) +$$
$$P(b_1b_4=b_1b_3) + P(b_1b_4=b_1b_4) + P(b_1b_4=b_2b_3) + P(b_1b_4=b_2b_4) +$$
$$P(b_2b_3=b_1b_3) + P(b_2b_3=b_1b_4) + P(b_2b_3=b_2b_3) + P(b_2b_3=b_2b_4) +$$
$$P(b_2b_4=b_1b_3) + P(b_2b_4=b_1b_4) + P(b_2b_4=b_2b_3) + P(b_2b_4=b_2b_4)]$$
$$= (1/16)(1 + 0 + 0 + 0 + 0 + 1 + 0 + 0 + 0 + 0 + 1 + 0 + 0 + 0 + 0 + 1) = 1/4.$$

Only one of four comparisons are expected to have genotypes at the "**b**" locus identical by descent. The average fraction of all loci with genotypes identical for pairs of full sibs is also one-fourth.

Dominance effects are defined as the interaction of two genes at one locus. Of the relatives shown in the summary table, only full sibs can have dominance effects contribute to likeness of pairs of relatives. Dominance effects occur when the value of b_ib_j is not the average value of b_i plus the average value of b_j. The dominance relationship between noninbred animals A and B, d_{AB}, can be found from the additive relationships among the parents of A and B as will be seen.

INBREEDING COEFFICIENT

The coefficient of inbreeding, F, is defined as the probability that two genes at one locus will be identical by descent when averaged over all loci, i.e., for an animal with one locus and genotype b_ib_j, $F = P(b_i=b_j)$. The two genes will be identical only if the parents have genes identical by descent. The expected frequency of two genes identical by descent at one locus is equal to the probability that each parent will contribute an identical gene, i.e., the probability of pairs of single genes being identical between the parents. Therefore, $F_p = (1/2)(\mathbf{a}_{sd})$ where **p**, s, and **d** refer to the progeny, sire, and dam, respectively, and $\mathbf{a}_{pp} = 1 + (1/2)(\mathbf{a}_{sd})$. F_p is the inbreeding coefficient which also corresponds to the fraction of loci having both genes identical by descent.

SOME USEFUL IDENTITIES IN WORKING WITH ADDITIVE AND DOMINANCE RELATIONSHIPS

If animals A and B have parents A_S, A_D and B_S, B_D, respectively, then usually

$$a_{AB} = (1/4)(a_{A_SB_S} + a_{A_SB_D} + a_{A_DB_S} + a_{A_DB_D}).$$

As shown in the appendix,

$$a_{AB} = (1/2)(a_{AB_S} + a_{AB_D}) \quad \text{if A is older than B.}$$

or equivalently $a_{AB} = (1/2)(a_{BA_S} + a_{BA_D})$ if B is older than A.

These equalities are the basis for computing additive relationships by the tabular method.

The dominance relationship can also be computed from the additive relationships among the parents even when the parents are inbred if the animals are themselves noninbred. As shown in the appendix,

$$d_{AB} = (1/4)({}^{a}A_{S}B_{S}\,{}^{a}A_{D}B_{D} + {}^{a}A_{S}B_{D}\,{}^{a}A_{D}B_{S}).$$

Also as just seen, the inbreeding coefficient for an animal is one-half the additive relationship between its parents,

$$F_A = (1/2)(a_{A_S A_D}) \; ; \; F_B = (1/2)(a_{B_S B_D}),$$

and an animal's additive relationship to itself is

$$a_{AA} = 1 + F_A = 1 + (1/2)(a_{A_S A_D}) \; ; \; a_{BB} = 1 + F_B = 1 + (1/2)(a_{B_S B_D}).$$

EXPANSION TO MORE THAN ONE LOCUS

The probability of a pair of nonallelic genes being alike in two individuals by descent is P(genes at first locus are identical by descent) × P(genes at the second locus are identical by descent). This pattern can be expanded to trios, etc., of nonallelic genes.

The probability of a particular combination of an allelic pair of genes (a genotype) and a gene at another locus being identical by descent in two individuals is P(the genotypes at one locus are identical) × P(genes at the other locus are identical).

The probability of a genotype at one locus and a genotype at another locus being common by descent in two individuals is P(first locus genotype is alike) ×P(other locus genotype is alike). However, these probabilities are equal. Thus, the probability of genotypes being common at two loci is P(genotype in common) squared or the square of the dominance relationship.

The expansion to higher order combinations can be done similarly.

To apply these principles, only two measures of relationship are needed: a_{ij}, the additive or **a** relationship between individuals **i** and **j** which is twice the fraction of single genes which are identical by descent (this will be the numerator of the coefficient of

relationship), and $\mathbf{d_{ij}}$, the probability that individuals **i** and **j** have a genotype at one locus (an allelic pair of genes) identical by descent (this is called the dominance or **d** relationship).

Many introductory texts describe how to calculate relationships by the method of tracing paths. That method is quick and easy for simple relationships and for few animals. A more powerful method described next is the tabular method (Cruden, 1949; Emik and Terrill, 1949).

TABULAR METHOD OF COMPUTING a_{ij} AND d_{ij}

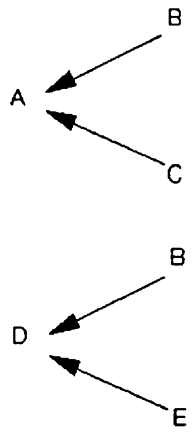
The easiest and safest method of computing additive relationships is the tabular method:

1. Determine which animals to include in the table. Include all animals after the oldest or base generation is chosen. Put them in order by date of birth, oldest first.
2. Write the names or numbers of the animals in order of birth across the top of the table (the columns) and along the side of the table (the rows) as shown in the example which follows.
3. Write above the number of the animals the numbers of their parents, if known.
4. Put a 1 in each of the diagonal cells of the table, such as row 1, column 1; row 2, column 2; etc. The one is the animal's basic relationship to itself unless it is inbred. For the base generation animals, enter their relationships to each other or assume them to be zero, and if known, add their inbreeding coefficients to the diagonals.

5. Begin at the diagonal of each row which now has a 1 in it. Add to this 1, one-half of the relationship between the animal's parents. This is the inbreeding coefficient which will often be zero. Compute the off-diagonal cells by rule 6.
6. Compute entries for each off-diagonal cell of row 1 according to the rule of 1/2 the entry for the first parent in this row plus 1/2 the entry for the second parent in the row. When the first row is finished, write the same values down the first column.
7. Continue as before for the next rows and columns until finished, always remembering to do a row at a time and to put the same values down the corresponding column before going to the next row.

Example

The following is an example of the $\mathbf{a}_{ij}$ and $\mathbf{d}_{ij}$ relationships for paternal half-sibs A and D.



				B-C	B-E
	B	C	E	A	D
B	1	0	0	1/2	1/2
C	0	1	0	1/2	0
E	0	0	1	0	1/2
A	1/2	1/2	0	1	1/4
D	1/2	0	1/2	1/4	1

$a_{BA} = (1/2)\ (a_{BB} + a_{BC}) = (1/2)\ (1 + 0) = 1/2$

$a_{BD} = (1/2)\ (a_{BB} + a_{BE}) = (1/2)\ (1 + 0) = 1/2$

$a_{CA} = (1/2)\ (a_{CB} + a_{CC}) = (1/2)\ (0 + 1) = 1/2$

$a_{CD} = (1/2)\ (a_{CB} + a_{CE}) = (1/2)\ (0 + 0) = 0$

$a_{EA} = (1/2)\ (a_{EB} + a_{EC}) = (1/2)\ (0 + 0) = 0$

$a_{ED} = (1/2)\ (a_{EB} + a_{EE}) = (1/2)\ (0 + 1) = 1/2$

$a_{AA} = 1 + (1/2)\ (a_{BC}) = 1 + (1/2)\ (0) = 1$

$a_{AD} = (1/2)\ (a_{AB} + a_{AE}) = (1/2)\ (1/2+0) = 1/4$

$a_{DD} = 1 + (1/2)\ (a_{BE}) = 1 + (1/2)\ (0) = 1$

The dominance relationship for non-inbred animals can be found from the additive relationships among the parents, e.g.,

$$d_{AD} = (1/4)\ (a_{BB} \times a_{CE} + a_{CB} \times a_{BE}) = (1/4)\ (1 \times 0 + 0 \times 0) = 0.$$

APPENDIX TO CHAPTER FIVE

PROOFS OF IDENTITIES FOR PROBABILITIES BY DESCENT

A. Probability of genes identical by descent: (Malécot, 1948)

1. Definition: Let the pair of animals, A and B, have genotypes $b_i b_j$ and $b_k b_\ell$ representing symbollically all loci, then

   P(random pair of genes identical) =

   . $\alpha_{AB} = \frac{1}{4}\,[P(i=k) + P(i=\ell) + P(j=k) + P(j=\ell)]$

2. Definition: The additive relationship, $a_{AB} = 2\,\alpha_{AB}$.

B. Probability of genotypes identical by descent:

1. Definition: Let A_S and A_D be the parents of A and B_S and B_D be the parents of B with genotypes $b_i b_j$ for A and $b_k b_\ell$ for B, then

   P(genotype identical) = $d_{AB} = P(b_i b_j = b_k b_\ell)$.

2. Computationally, $d_{AB} = 1/4\,(a_{A_S B_S} \times a_{A_D B_D} + a_{A_S B_D} \times a_{A_D B_S})$ for non-inbred animals:

   $P(b_i b_j = b_k b_\ell)$ = P(A_S contributes b_i to A and B_S contributes $b_{i=k}$ to B) ×
   P(A_D contributes b_j to A and B_A contributes $b_{j=\ell}$ to B)
   + P(A_S contributes b_i to A and B_D contributes $b_{i=k}$ to B) ×
   P(A_D contributes b_j to A and B_S contributes $b_{j=\ell}$ to B)

But, P(A_S contributes b_i to A and B_S contributes b_i to B)

= P(genes identical by descent for A_S and B_S)) = $\alpha_{A_S B_S}$.

Similarly for the other probabilities.

Thus

$$P(b_ib_j = b_kb_\ell) = \alpha_{A_SB_S} \times \alpha_{A_DB_D} + \alpha_{A_SB_D} \times \alpha_{A_DB_S}$$

$$= (1/4)\left[a_{A_SB_S} \times a_{A_DB_D} + a_{A_SB_D} \times a_{A_DB_S}\right]$$

The four additive relationships can be found from the relationship table.

C. The inbreeding coefficient, F_A, is the fraction of loci with genes identical by descent for animal A.

By definition, F of the loci of A have $\alpha = 1$ and 1 - F of the loci have $\alpha = 1/2$,

i.e., F of the loci are of the form b_ib_i with $\alpha = 1$ and 1-F of the loci are of the form b_ib_j with $\alpha = 1/2$.

1. Thus the average fraction of genes identical for A with itself is:

   α_{AA} = P(genes identical) = (F)(1) + (1-F) (1/2) = 1/2 + (1/2)F and

   $a_{AA} = 2\alpha_{AA} = 1 + F$, that is, the numerator or additive relationship of an individual to itself is 1 + F.

2. If S is the sire of A and D is the dam of A, then $F_A = (1/2)\, a_{SD}$.

   Let the genotypes be b_ib_j for A, b_kb_ℓ for S, and b_mb_n for D.

   By definition $\alpha_{AA} = (1/4)\left[P(i=i) + P(i=j) + P(j=i) + P(j=j)\right]$

   $= 1/2 + (1/2)\, P(i=j)$

Thus $F_A = P(i=j)$.

But P(i=j) is α_{SD} since b_i must come from one parent and b_j from the other,

i.e., $P(i=j) = (1/4)\left[P(k=m) + P(k=n) + P(\ell=m) + P(\ell=n)\right]$

$= \alpha_{SD}$

Therefore, $F_A = \alpha_{SD} = (1/2)\, a_{SD}$.

D. The tabular method of computing relationships depends on the fact that if B has parents B_S and B_D, then $a_{AB} = (1/2)\ (a_{AB_S} + a_{AB_D})$.

Let the symbolic genotypes be b_ib_j for A, b_kb_ℓ for B_S and b_mb_n for B_D.

The possible and equally likely genotypes of B are:

B_1, b_kb_m;

B_2, b_kb_n;

B_3, $b_\ell b_m$; and

B_4, $b_\ell b_n$.

By previous definition,

$$\alpha_{AB} = \text{average of } \alpha_{AB_1},\ \alpha_{AB_2},\ \alpha_{AB_3},\ \alpha_{AB_4} \quad \text{so that:}$$

$$\begin{aligned}\alpha_{AB} = (1/4)\ \{ &(1/4)[P(i=k) + P(i=m) + P(j=k) + P(j=m)] \\ &+ (1/4)\ [P(i=k) + P(i=n) + P(j=k) + P(j=n)] \\ &+ (1/4)\ [P(i=\ell) + P(i=m) + P(j=\ell) + P(j=m)] \\ &+ (1/4)\ [P(i=\ell) + P(i=n) + P(j=\ell) + P(j=n)]\}\end{aligned}$$

After combining and rearranging:

$$\alpha_{AB} = (1/8)[P(i=k) + P(i=\ell) + P(i=m) + P(i=n) + P(j=k) + P(j=\ell) + P(j=m) + P(j=n)]$$

But $\alpha_{AB_S} = (1/4)[P(i=k) + P(i=\ell) + P(j=k) + P(j=\ell)]$ and

$\alpha_{AB_D} = (1/4)[P(i=m) + P(i=n) + P(j=m) + P(j=n)]$.

Thus $\alpha_{AB} = (1/2)\ (\alpha_{AB_S} + \alpha_{AB_D})$ and $a_{AB} = (1/2)\ (a_{AB_S} + a_{AB_D})$.

CHAPTER 6

GENETIC VALUES AND GENETIC COVARIANCES

Quantitative geneticists have followed the nomenclature of statistics and the logic of effects of different types of gene combinations to define several kinds of genetic effects. These types include single gene effects which generally are most important and effects of gene combinations such as a gene pair at one locus (the genotype) and a gene pair with the genes at different loci. The potential number of combinations of more genotypes and genes at separate loci is nearly infinite. In introductory texts the combinations are put into two groups: 1) single gene effects which together sum to breeding value or additive genetic value and 2) all other genetic effects which are called epistatic or interaction effects. In this chapter the theoretical partition of genetic effects into as many combinations as might be needed will be discussed, even though one type of effect is usually of primary importance, the additive gene effects. Two other kinds of effects have received some practical attention, dominance (the gene pair at one locus) effects, and additive by additive (gene pairs with the genes at different loci) effects. The definitions also lead directly to describing the genetic covariance between records of relatives with specified additive and dominance relationships.

DEFINITION OF GENETIC VALUES

An additive gene effect is defined as the average replacement value of that gene, i.e., if that gene replaces the average gene, the change in value is the additive genetic effect of

that gene. Thus, if two of that gene are added, the change in value will be twice the additive effect of adding one gene. The sum over all loci of all additive genetic effects is the additive genetic value, G_A, of the animal.

A dominance genetic effect is defined as the average replacement value of a particular gene pair at one locus as a difference from the additive genetic effects. The sum over all loci of all dominance genetic effects is the dominance genetic value, G_D, of the animal.

An additive by additive genetic effect is defined as the average replacement value of a pair of non-allelic genes--the specific effect of a gene from one locus and a gene from another locus as a difference from the additive genetic effects of the genes. The additive by additive genetic value, G_{AA}, of an animal is the sum of all specific effects of non-allelic gene pairs.

An additive by dominance gene effect is defined as the average replacement value of a gene at one locus and a gene pair (genotype) at another locus as a difference from the additive, dominance, and additive by additive genetic effects. The sum of all such effects is the additive by dominance genetic value, G_{AD}, of an animal.

Similarly, higher order genetic effects can be defined, e.g., additive by additive by additive and additive by dominance by dominance. These different types of genetic effects are defined to be independent and to have average values of zero in an unselected population.

The total genetic value of an animal is the sum of the various genetic values:

$$G = G_A + G_D + G_{AA} + G_{AD} + G_{AAA} + \cdots$$

If these values could be measured separately, variances for each could be computed as for any other variable. Whether or not they can be measured, a variance can be hypothesized for each kind of genetic value. In fact, since the various genetic values are defined to be independent, total genetic variance is the sum of the variances of the component genetic values:

$$\sigma^2_G = \sigma^2_{G_A} + \sigma^2_{G_D} + \sigma^2_{G_{AA}} + \sigma^2_{G_{AD}} + \sigma^2_{G_{AAA}} + \cdots$$

A simpler but less symbolic notation for the components of genetic variance is based on σ^2_{ij} where i refers to the number of single nonallelic genes and j refers to the number of allelic pairs (genotypes) contributing to the genetic effect. This notation is summarized as follows:

Gene action	Contribution to genetic variation		
sum of effects of	symbols		jargon
single genes: a_1, a_3, b_1, b_{10}, etc.	$\sigma^2_{G_A}$	σ^2_{10}	additive genetic variance
allelic pairs: a_1a_2, c_1c_5, etc.	$\sigma^2_{G_D}$	σ^2_{01}	dominance genetic variance
non-allelic pairs: a_1b_1, a_2c_5, etc.	$\sigma^2_{G_{AA}}$	σ^2_{20}	additive by additive
single genes and allelic pairs: $a_1b_1b_2$, $c_1d_5d_6$, etc.	$\sigma^2_{G_{AD}}$	σ^2_{11}	additive by dominance
two allelic pairs: $a_1a_3b_4b_6$, $c_1c_2b_2b_3$, etc.	$\sigma^2_{G_{DD}}$	σ^2_{02}	dominance by dominance
in general		σ^2_{ij}	where i refers to number of nonallelic genes acting together with j allelic pairs

Total genetic variance can then be written as:

$$\sigma^2_G = \sum_{i+j>0} \sum \sigma^2_{ij} = \sigma^2_{10} + \sigma^2_{01} + \sigma^2_{20} + \sigma^2_{11} + \sigma^2_{02} + \cdots$$

GENE EFFECTS IN COMMON BY DESCENT

Since there are two additive genetic effects at each locus, the fraction of additive gene effects in common for relatives A and B is the additive or numerator relationship, a_{AB}, which equals twice the probability that a random gene from animal A is identical by descent to a random gene from relative B at a single locus. The appendix to this chapter contains a more mathematical explanation for using a_{AB} rather than $a_{AB}/2$ to describe additive effects in common by descent.

The fraction of dominance effects in common will be d_{AB} which equals the probability of genotypes identical by descent or equivalently the fraction of loci with identical genotypes for relatives A and B.

Similarly, a^2_{AB} is the fraction of additive by additive genetic effects in common and $a_{AB}d_{AB}$ is the fraction of additive by dominance genetic effects identical by descent.

In general, $(a_{AB})^i(d_{AB})^j$ gives the fraction of genetic effects in common by descent due to **i** non-allelic genes acting together with **j** allelic pairs (genotypes).

GENETIC COVARIANCES BETWEEN RELATIVES

Genetic covariance between relatives depends on the fraction of different kinds of genetic effects which are common by descent. In fact, covariance due to additive gene effects in common is $a_{AB}\sigma^2_{10}$ -- the product of the fraction of additive effects in common and additive genetic variance. Covariance due to common dominance effects is $d_{AB}\sigma^2_{01}$; that due to additive by additive effects is $a^2_{AB}\sigma^2_{20}$ and that due to additive by dominance effects is $a_{AB}d_{AB}\sigma^2_{11}$. These and others are summarized as follows:

Contribution to genetic covariance between individuals.

Genetic components	Contribution to covariance between individuals A and B
σ^2_{10}	$(a_{AB})^1 \quad \sigma^2_{10}$
σ^2_{01}	$(d_{AB})^1 \quad \sigma^2_{01}$
σ^2_{20}	$(a_{AB})^2 \quad \sigma^2_{20}$
σ^2_{11}	$(a_{AB})^1(d_{AB})^1 \quad \sigma^2_{11}$
σ^2_{02}	$(d_{AB})^2 \quad \sigma^2_{02}$
σ^2_{12}	$(a_{AB})^1(d_{AB})^2 \quad \sigma^2_{12}$
.	.
.	.
.	.
σ^2_{ij}	$(a_{AB})^i(d_{AB})^j \quad \sigma^2_{ij}$

for i = 0, . . ., n; j = 0, . . ., n with n loci, also i + j must be > 0 and i + j ≤ n

The total genetic covariance is the sum of the parts, that is,

$$\sigma_{G_A G_B} = a_{AB}\sigma^2_{10} + d_{AB}\sigma^2_{01} + a^2_{AB}\sigma^2_{20} + a_{AB}d_{AB}\sigma^2_{11} + \cdots .$$

In summation notation the total genetic covariance can be written as

$$\sigma_{G_A G_B} = \sum_{i+j>0} \sum (a_{AB})^i (d_{AB})^j \sigma^2_{ij} .$$

The subscripts of the genetic variance components correspond to the superscripts of the additive and dominance relationships. When j = 0, $(d_{AB})^j = 1$ for any d_{AB} and when $d_{AB} = 0$, $(d_{AB})^0 = 1$ but $(d_{AB})^1 = 0$, etc. These simplifications are illustrated in the coefficients in the column for contribution to covariance between individuals. An important point is that as **i** increases, the coefficients of the higher order genetic components of variance decrease. Thus even if σ^2_{i0} is large, the contribution to likeness by that component, $(a_{AB})^i\sigma^2_{i0}$, will be small if **i** is very large. For example, with $a_{AB} = 1/4$

and i = 3, a^3_{AB} = 1/64 is the coefficient for variance due to additive by additive by additive effects as compared to the coefficient of 1/4 for variance due to additive genetic variance.

EXPECTED VALUES TO SHOW CONTRIBUTION TO GENETIC COVARIANCE

As shown in Chapter 1, the average or expected value of the product of two variables, x_i and y_i, is written $E(x_iy_i)$. If $E(x_i) = 0$ is the average of variable X, then $E(x_iy_i) = \sigma_{xy}$. Similarly, $E(x_i^2) = \sigma_x^2$, $E(kx_i)^2 = k^2\sigma_x^2$ where k is a constant and $E(k_xx_ik_yy_i) = k_xk_y\sigma_{xy}$.

This principle will be applied to the example of genetic covariance between relatives X and Y for only three kinds of genetic effects but this example will illustrate how the overall genetic covariance between relatives is determined.

Let $G_X = G_{A_X} + G_{D_X} + G_{AA_X}$. If Y is related to X, then a fraction of these gene effects also appear in G_Y. Then write

$$G_Y = G_{A_Y} + G_{D_Y} + G_{AA_Y}$$

$$G_Y = a_{XY}G_{A_X} + \text{other } G_{A_Y} + d_{XY}G_{D_X} + \text{other } G_{D_Y} + a^2_{XY}G_{AA_X} + \text{other } G_{AA_Y}.$$

The other genetic effects are due to genes from other sources and Mendelian sampling and are independent of the effects in common with G_X.

Since the genetic effects are defined to be independent with zero means, then

$$\sigma_{G_XG_Y} = E(G_XG_Y).$$

Next, substitute for G_X and G_Y, expand, and take expectations of the parts:

$$\begin{aligned}\sigma_{G_X G_Y} = {} & E[(G_{A_X})(a_{XY}G_{A_X})] + E[(G_{A_X})(\text{other } G_{A_Y})] + E[(G_{A_X})(d_{XY}G_{D_X})] + \\ & E[(G_{A_X})(\text{other } G_{D_Y})] + E[(G_{A_X})(a^2_{XY}G_{AA_X})] + E[(G_{A_X})(\text{other } G_{AA_Y})] + \\ & E[(G_{D_X})(a_{XY}G_{A_X})] + E[(G_{D_X})(\text{other } G_{A_Y})] + E[(G_{D_X})(d_{XY}G_{D_X})] + \\ & E[(G_{D_X})(\text{other } G_{D_Y})] + E[(G_{D_X})(a^2_{XY}G_{AA_X})] + E[(G_{D_X})(\text{other } G_{AA_Y})] + \\ & E[(G_{AA_X})(a_{XY}G_{A_X})] + E[(G_{AA_X})(\text{other } G_{A_Y})] + E[(G_{AA_X})(d_{XY}G_{D_X})] + \\ & E[(G_{AA_X})(\text{other } G_{D_Y})] + E[(G_{AA_X})(a^2_{XY}G_{AA_X})] + E[(G_{AA_X})(\text{other } G_{AA_Y})] .\end{aligned}$$

After factoring constants outside the expected value operator, then according to the rules for expected values:

$$\begin{aligned}\sigma_{G_X G_Y} = {} & a_{XY}\sigma^2_{G_A} + 0 + 0 + 0 + 0 + 0 + 0 + 0 + d_{XY}\,\sigma^2_{G_D} + \\ & 0 + 0 + 0 + 0 + 0 + 0 + 0 + a^2_{XY}\sigma^2_{G_{AA}} + 0 .\end{aligned}$$

The zero terms come from independence of genetic effects and the lack of genetic effects in common between terms such as (G_{A_X}) and $(\text{other } G_{A_Y})$.

Example one: The contribution of all genetic components up to second order ($i + j = 2$) interaction components to the likeness between records of a parent (X) and its progeny (Y).

The relationships are: $a_{XY} = \frac{1}{2}$ and $d_{XY} = 0$.

Therefore:

$$\begin{aligned}\sigma_{G_X G_Y} = {} & (\tfrac{1}{2})^1(0)^0\,\sigma^2_{10} + (\tfrac{1}{2})^0(0)^1\,\sigma^2_{01} + (\tfrac{1}{2})^2(0)^0\,\sigma^2_{20} + \\ & (\tfrac{1}{2})^0(0)^2\,\sigma^2_{02} + (\tfrac{1}{2})^1(0)^1\,\sigma^2_{11} .\end{aligned}$$

Thus:

$$\sigma_{G_X G_Y} = (\tfrac{1}{2})\,\sigma^2_{10} + (\tfrac{1}{2})^2\,\sigma^2_{20} .$$

(Note that $(0)^0 = (N)^0 = 1$ for any number (N), but that $(0)^N = 0$ for $N > 0$.)

Example Two: The genetic covariance between full sibs, X and Y.

Because $a_{XY} = \frac{1}{2}$ and $d_{XY} = \frac{1}{4}$,

then

$$\sigma_{G_X G_Y} = (\tfrac{1}{2})\,\sigma^2_{10} + (\tfrac{1}{4})\,\sigma^2_{01} + (\tfrac{1}{2})^2\,\sigma^2_{20} + (\tfrac{1}{4})^2\,\sigma^2_{02} + (\tfrac{1}{2})(\tfrac{1}{4})\,\sigma^2_{11}\,.$$

Although full sib pairs and parent-progeny pairs have the same additive relationship the likeness (genetic covariance) will be greater between full sib pairs than parent-progeny pairs if dominance effects, dominance by dominance effects, and additive by dominance effects contribute to genetic variation.

ESTIMATION OF GENETIC VARIANCES

These two examples also indicate how the components of genetic variance may be estimated. Covariances between pairs of relatives are computed and equated to their theoretical composition. In general, as many covariances as theoretical components are necessary. In the above two examples, only two components could be estimated but not σ^2_{10} and σ^2_{20} together since both the parent-progeny and full sib covariances have the same expectation for those components. Usually σ^2_{10} and σ^2_{01} would be estimated for this case. Note that the other components usually must be assumed to be zero.

For example, suppose that Cov(full sib one, full sib two) = 50, and also that Cov(parent-progeny) = 40. Assume $\sigma^2_{20} = \sigma^2_{02} = \sigma^2_{11} = 0$.

Then,

$$50 = (\tfrac{1}{2})\,\sigma^2_{10} + (\tfrac{1}{4})\,\sigma^2_{01}$$

$$40 = (\tfrac{1}{2})\,\sigma^2_{10}$$

Thus, estimates are $\hat{\sigma}^2_{10} = 80$ and $\hat{\sigma}^2_{01} = 40$.

In general, for a random mating population, the additive fraction of genetic variance, σ^2_{10}, is about all that can be used for selection gains. Selection for gene combinations is ineffective because the contribution to descendants drops by a_{AB} with each generation. The usual goal is to select for additive merit--the part that contributes σ^2_{10} to genetic variance and the most to covariances between relatives.

DEFINITION OF HERITABILITY

Heritability in the "broad sense" is defined as $\sigma^2_G/(\sigma^2_G+\sigma^2_E)$ where σ^2_G is the total genetic variance, $\Sigma\Sigma_{ij}\ \sigma^2_{ij}$, and σ^2_E is the variance due to non-genetic effects (environmental effects).

Heritability in the "narrow sense" is defined as $\sigma^2_{10}/(\sigma^2_G+\sigma^2_E)$ where σ^2_{10} is the additive genetic variance and $\sigma^2_G + \sigma^2_E$ is the total or phenotypic variance which is the total genetic variance plus the environmental variance. This form of heritability, sometimes called additive heritability, will be used again and again when methods of selection for additive genetic value are discussed.

APPENDIX TO CHAPTER SIX

WHY a_{AB} DESCRIBES GENETIC COVARIANCE RATHER THAN $a_{AB}/2$

1. Additive Genetic Variance, σ^2_{10}

Consider one locus only.

Let $\alpha_i + \alpha_j = G_A$ G_A is additive genetic value of animal A due to effects α_i and α_j of genes a_i and a_j

$\alpha_{i'} + \alpha_{j'} = G_B$ G_B is additive genetic value of animal B due to effects $\alpha_{i'}$ and $\alpha_{j'}$ of genes $a_{i'}$ and $a_{j'}$.

Then $COV\ (G_A,G_B) = E(\alpha_i\alpha_{i'} + \alpha_i\alpha_{j'} + \alpha_j\alpha_{i'} + \alpha_j\alpha_{j'})$

But $E(\alpha^2_m) = \alpha^2$ for all m and $E(\alpha_m\alpha_{m'}) = 0$

for all $m \neq m'$.

Thus $COV\ (G_A,G_B) = \alpha^2[P(i=i') + P(i=j') + P(j=i') + P(j=j')]$

$= \alpha^2$ [4P (random genes of A and B are identical by descent)]

$= \alpha^2\ [4\ (1/2)a_{AB}]$

$= \alpha^2\ [2a_{AB}]$,

but $\sigma^2_{10} = E[G^2_A] = E[\alpha_i + \alpha_j]^2$

$= E[\alpha^2_i + \alpha^2_j + 2\alpha_i\alpha_j]$

$= 2\alpha^2 + 0$ since $E(\alpha_i) = 0$

and $E(\alpha_i\alpha_j) = 0$ unless inbred.

Thus $\dfrac{\sigma^2_{10}}{2} = \alpha^2$ and

therefore $COV\ (G_A,G_B) = \dfrac{\sigma^2_{10}}{2}\ (2a_{AB}) = a_{AB}\sigma^2_{10}$. This procedure may be extended to many loci.

2. Additive by Additive Genetic Variance, σ^2_{20}

Consider the minimum of two loci and let $(\alpha\beta)_{mn}$ be the additive by additive effect of the $m^{\underline{th}}$ gene from the "**a**" locus and $n^{\underline{th}}$ gene from the "**b**" locus. Let the additive values of animals A and B be

$G_{10,A} = \alpha_i + \alpha_j + \beta_k + \beta_\ell$ and

$G_{10,B} = \alpha_{i'} + \alpha_{j'} + \beta_{k'} + \beta_{\ell'}$.

Then let the corresponding additive by additive effects be

$G_{20,A} = (\alpha\beta)_{ik} + (\alpha\beta)_{i\ell} + (\alpha\beta)_{jk} + (\alpha\beta)_{j\ell}$ and

$G_{20,B} = (\alpha\beta)_{i'k'} + (\alpha\beta)_{i'\ell'} + (\alpha\beta)_{j'k'} + (\alpha\beta)_{j'\ell'}$.

$$\begin{aligned}\text{Then } COV_{20}(A,B) = (\alpha\beta)^2[&P(i=i')P(k=k')+P(i=i')P(k=\ell')+P(i=j')P(k=k')+P(i=j')P(k=\ell')\\ &+P(i=i')P(\ell=k')+P(i=i')P(\ell=\ell')+P(i=j')P(\ell=k')+P(i=j')P(\ell=\ell')\\ &+P(j=i')P(k=k')+P(j=i')P(k=\ell')+P(j=j')P(k=k')+P(j=j')P(k=\ell')\\ &+P(j=i')P(\ell=k')+P(j=i')P(\ell=\ell')+P(j=j')P(\ell=k')+P(j=j')P(\ell=\ell')]\end{aligned}$$

$$\begin{aligned}&= (\alpha\beta)^2[P(i=i')+P(i=j')+P(j=i')+P(j=j')]\\ &\quad x[P(k=k')+P(k=\ell')+P(\ell=k')+P(\ell=\ell')]\\ &= (\alpha\beta)^2[4P(\text{genes identical})x4P(\text{genes identical})]\\ &= (\alpha\beta)^2[(2a_{AB})(2a_{AB})]\end{aligned}$$

But $\sigma^2_{20} = E[(G_{20,A})^2] = 4(\alpha\beta)^2$ so that $\sigma^2_{20}/4 = (\alpha\beta)^2$

Therefore $COV_{20}(A,B) = \frac{\sigma^2_{20}}{4}(4a^2_{AB}) = a^2_{AB}\sigma^2_{20}$ when all terms are evaluated.

CHAPTER 7

THE SELECTION INDEX

The basic problem in obtaining improvement through breeding is to choose animals that have the greatest genetic value to be parents of the next generation. The simplified model for a record, P_i, on animal i poses the problem:

$$P_i = \mu + G_i + E_i ,$$

where μ is the population mean, a constant, which may represent other fixed factors that influence P_i; G_i is the effect on P_i due to the animal's complete genotype, and E_i is the effect of the environment on P_i and is the effect that masks the evaluation of G_i. As was demonstrated earlier, only additive genetic effects have much chance of being transmitted over many generations. However, often G_i can be safely assumed to be due only to additive genetic effects.

The problem is to maximize the average of G of the selected group, μ_{Gs}, where μ_G is the average G of the total group, i.e.,

$$\text{MAXIMIZE } [\Delta G = \mu_{Gs} - \mu_G].$$

Genetic improvement per year under normality and other assumptions, as will be derived later, is:

$$\Delta G/yr = (r_{TI} D \sigma_G)/L ,$$

where ΔG is the genetic improvement per generation; r_{TI} is the correlation between the true additive genetic value and I, the index prediction of it; D is a factor related to selection intensity (value of 0 with no selection and a value of about 3 for selection of the top one-half percent); σ_G is the genetic standard deviation, and L is the generation interval in years defined as the average number of years between birth of parents and the birth of replacement offspring. The four parts of the key equation for genetic improvement will be discussed separately.

What is a selection index estimate of genetic value? This question is, perhaps, best answered by an example.

Suppose several animals each have three relatives with records; (X_1, X_2, and X_3). Relatives are known to have genetic effects in common by descent. Thus, the record of each relative should tell something about the genetic value of the animal being evaluated. A logical way to put the information together is to weight each record by its relative importance, i.e., estimate G as $I = b_1X_1 + b_2X_2 + b_3X_3$, where the **b**'s are the appropriate weights and the X's are known records of the three relatives. The selection index prediction of true genetic value is I. The records are adjusted for any fixed factors such as μ, i.e., $X_i = P_i - \mu$.

WHAT SHOULD THE WEIGHTS (b's) BE?

Some desirable properties of the index to predict some true value, T, should be:

1. To minimize errors of prediction which is the average or expected squared difference between T and its predictor, I, i.e., **MINIMIZE** $E(T-I)^2$.

2. To maximize r_{TI}, the correlation between true value and prediction of true value; this correlation is often called accuracy of prediction of T.
3. To maximize the probability of correctly ranking the animals, and
4. To maximize average true value of the selected group.

The selection index procedure which will be described satisfies properties 1 and 2 and satisfies properties 3 and 4 if the records of relatives, the X's, and T, the true value, follow a multivariate normal distribution. These procedures were developed from work by Sewall Wright, Jay Lush, and C. R. Henderson. Henderson proved many of the properties. Most of the development that follows was taught for many years by C. R. Henderson at Cornell University, beginning in 1948.

METHOD OF FINDING b's

The general linear index is $I = b_1X_1 + b_2X_2 + \cdots + b_NX_N$ for predicting some true value, T, which often is, but is not necessarily, additive genetic value. The goal is to maximize r_{TI}. Maximizing the logarithm of r_{TI}, log (r_{TI}), is equivalent to maximizing r_{TI} but is easier to accomplish. Note that:

$$\log (r_{TI}) = \log (\sigma_{TI}) - (1/2) \log (\sigma^2_T) - (1/2) \log (\sigma^2_I).$$

The rules for finding variances and covariances of linear functions (see Chapter 1) will give σ_{TI} and σ^2_I in terms of the unknown **b**'s and known variances and covariances. Note that σ^2_T is a constant and does not contribute to the equations.

$$\sigma_{TI} = b_1\sigma_{TX_1} + b_2\sigma_{TX_2} + \cdots + b_N\sigma_{TX_N}, \text{ and}$$

$$\sigma^2_I = b_1^2\sigma^2_{X_1} + 2b_1b_2\sigma_{X_1X_2} + \cdots + 2b_1b_N\sigma_{X_1X_N} + b_2^2\sigma^2_{X_2} + 2b_2b_3\sigma_{X_2X_3} + \cdots + b_N^2\sigma^2_{X_N}.$$

These expressions are then substituted into log (r_{TI}), and the partial derivatives of log (r_{TI}) with respect to each of the **b**'s are set equal to zero, i.e.,

$$\frac{\delta \log(r_{TI})}{\delta b_1} = \frac{\sigma_{X_1 T}}{\sigma_{TI}} - \frac{b_1\sigma^2_{X_1} + b_2\sigma_{X_1X_2} + \cdots + b_N\sigma_{X_1X_N}}{\sigma^2_I} = 0$$

$$\frac{\delta \log(r_{TI})}{\delta b_2} = \frac{\sigma_{X_2 T}}{\sigma_{TI}} - \frac{b_1\sigma_{X_1X_2} + b_2\sigma^2_{X_2} + \cdots + b_N\sigma_{X_2X_N}}{\sigma^2_I} = 0$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\frac{\delta \log(r_{TI})}{\delta b_N} = \frac{\sigma_{X_N T}}{\sigma_{TI}} - \frac{b_1\sigma_{X_1X_N} + b_2\sigma_{X_2X_N} + \cdots + b_N\sigma^2_{X_N}}{\sigma^2_I} = 0$$

Rearrangement of these equations gives the selection index equations (except for a constant, $k = \sigma^2_I/\sigma_{TI}$, on the right hand sides of the equations) which define the unknown selection index weights, the **b**'s:

$$b_1\sigma^2_{X_1} + b_2\sigma_{X_1X_2} + b_3\sigma_{X_1X_3} + \cdots + b_N\sigma_{X_1X_N} = k\sigma_{X_1T}$$

$$b_1\sigma_{X_2X_1} + b_2\sigma^2_{X_2} + b_3\sigma_{X_2X_3} + \cdots + b_N\sigma_{X_2X_N} = k\sigma_{X_2T}$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$b_1\sigma_{X_NX_1} + b_2\sigma_{X_NX_2} + b_3\sigma_{X_NX_3} + \cdots + b_N\sigma^2_{X_N} = k\sigma_{X_NT}$$

Some important points to notice about these equations are:

1. The constant, $\mathbf{k} = \sigma_I^2/\sigma_{TI}$, will not change the relative sizes of the **b**'s or the r_{TI} so **k** can be set equal to 1, which as will be shown, will result in the same **b**'s that minimize squared prediction error. In fact, when squared prediction error is minimized, $\sigma_I^2 = \sigma_{TI}$.
2. The equations are symmetrical, i.e., the coefficients of the unknown **b**'s are the same in each column as the corresponding row. See, for example, the coefficients (the covariances) in row 1 and in column 1.
3. The equations are similar to multiple regression equations except that the true variances and covariances are assumed known and replace the sums of squares and products used in multiple regression.
4. If squared prediction error, $E(T-I)^2$, is minimized, the same equations are found except that the constant, σ_I^2/σ_{TI}, is not a multiplier of the right-hand sides of the equations.

Average squared prediction error is

$$E[(T-I)^2] = \sigma_T^2 + \sigma_I^2 - 2\sigma_{TI} + \mu_T^2 + \mu_I^2 - 2\mu_T\mu_I ,$$

The constants μ_T and μ_I will not change differences in the I's. Thus, σ_T^2, σ_I^2, and σ_{TI} can be expressed in terms of linear functions of the **b**'s as for maximization of r_{TI}. Usually μ_T and μ_I are assumed to equal zero. The partial derivatives of $\sigma_T^2 + \sigma_I^2 - 2\sigma_{TI}$ with respect to $\mathbf{b}_1, \mathbf{b}_2 \cdots, \mathbf{b}_N$ equated to zero provide the following equations which define the **b**'s which minimize prediction error squared and also maximize r_{TI}:

$$\frac{\delta(\sigma_I^2 + \sigma_T^2 - 2\sigma_{TI})}{\delta b_1} = 0: \quad b_1\sigma_{X_1}^2 + b_2\sigma_{X_1X_2} + \cdots + b_N\sigma_{X_1X_N} = \sigma_{X_1T}$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$\frac{\delta(\sigma_I^2 + \sigma_T^2 - 2\sigma_{TI})}{\delta b_N} = 0: \quad b_1\sigma_{X_NX_1} + b_2\sigma_{X_NX_2} + \cdots + b_N\sigma_{X_N}^2 = \sigma_{X_NT}$$

These equations are the same as for maximizing r_{TI} when σ_I^2/σ_{TI} is set equal to unity. In this derivation, $\sigma_I^2 = \sigma_{TI}$ automatically as shown in the appendix to this chapter.

OTHER PROPERTIES OF THE SELECTION INDEX

1. The correlation between the index and true value is:

$$r_{TI} = \sqrt{\Sigma b_i\sigma_{X_iT}/\sigma_T^2} = \sqrt{(b_1\sigma_{X_1T} + b_2\sigma_{X_2T} + \cdots)/\sigma_T^2} \ .$$

The rules for expected values show:

$$\sigma_{TI} = \Sigma b_i\sigma_{X_iT} \quad \text{so that} \quad r_{TI} = \sqrt{\sigma_{TI}/\sigma_T^2} \ .$$

If an index is not the selection index, the definitional form of the correlation must be used to obtain accuracy:

$$r_{TI} = \sigma_{TI}/\sqrt{\sigma_T^2\sigma_I^2} \ ,$$

where σ_{TI} and σ_I^2 can be calculated using expected values.

If the index is the selection index, the definitional form of the correlation reduces to:

$$\sqrt{\sigma_{TI}/\sigma_T^2} \quad \text{because} \quad \sigma_I^2 = \sigma_{TI}.$$

2. Because $I = \Sigma b_i X_i$ and the X_i are variables, then the index values will also be variable. In fact, if I is the selection index:

$$\sigma_I^2 = r_{TI}^2 \sigma_T^2 .$$

This expression shows that σ_I^2 corresponds to variation in T that is accounted for by I.

When I is not the selection index:

$$\sigma_I^2 = E(I^2) * r_{TI}^2 \sigma_T^2 .$$

3. The variance of prediction errors (average squared difference of T from I) is:

$$V(T\text{-}I) = E[(T\text{-}I)^2] = (1\text{-}r_{TI}^2)\sigma_T^2 .$$

This expression corresponds to the variation in T not accounted for by I.

When I is not the selection index, the variance of prediction errors must be calculated from expected values:

$$E[(T\text{-}I)^2] = E(T^2) + E(I^2) - 2E(TI) * (1\text{-}r_{TI}^2)\sigma_T^2 .$$

4. The average of true values for animals with index value I_o is:

$$E(T \mid I = I_o) = I_o .$$

With this property, the selection index procedure is unbiased.

5. Intuitively, animals with the same index value would be expected to have different true values. In fact, the variance of true values for animals with the same index value, I_o, is:

$$V(T \mid I = I_o) = (1\text{-}r_{TI}^2)\sigma_T^2 .$$

These properties will be used later to make probability statements about the true value of an animal with a certain index value. If I is not the selection index, r_{TI} must be calculated from E(TI), $E(I^2)$, and σ_T^2.

APPENDIX TO CHAPTER SEVEN

DERIVATION OF SELECTION INDEX WITH MATRIX ALGEBRA

Let **x** be the information vector:

$$\mathbf{x} = (x_1 \cdots x_N)'.$$

T may be a true value for a trait, or, e.g., T may be a combination of a vector of genetic values, **g**, for several traits weighted by a vector of economic values, **v**, i.e.:

$$T = \mathbf{v}'\mathbf{g}.$$

T is to be predicted from a linear function of **x**; that is, each $\mathbf{x}_i$ is weighted by some factor $\mathbf{b}_i$ so that $\hat{T} = I = \mathbf{b}'\mathbf{x}$. With no loss of generality, the $\mathbf{x}_i$ can be assumed to have zero means, i.e., have been adjusted for fixed effects such as μ. The variance-covariance matrix of **x** is $E(\mathbf{x}\mathbf{x}') = P$.

Then:

$$\sigma_I^2 = E[\mathbf{b}'\mathbf{x}\mathbf{x}'\mathbf{b}] = \mathbf{b}'E[\mathbf{x}\mathbf{x}']\mathbf{b} = \mathbf{b}P\mathbf{b}' .$$

With T, the scalar variable to be predicted, $\sigma_{TI} = E[T\mathbf{b}'\mathbf{x}] = \mathbf{b}'E[T\mathbf{x}] = \mathbf{b}'\mathbf{c}$ where **c** is the vector of covariances between the $\mathbf{x}_i$ and T, e.g.:

$$(\sigma_{X_1T} \quad \sigma_{X_2T} \cdots \sigma_{X_NT})' .$$

Squared prediction error is:

$$\sigma_T^2 + \sigma_I^2 - 2\sigma_{TI} = \sigma_T^2 + \mathbf{b}'P\mathbf{b} - 2\mathbf{b}'\mathbf{c}.$$

To minimize squared prediction error, partial derivatives are taken with respect to **b** and equated to zero with rules for derivatives of matrices and vectors (Searle, 1982):

$$\frac{\delta(\sigma_T^2 + \mathbf{b}'\mathbf{P}\mathbf{b} - 2\mathbf{b}'\mathbf{c})}{\delta \mathbf{b}} = 0, \text{ where by parts;}$$

$$\frac{\delta(\sigma_T^2)}{\delta \mathbf{b}} = 0, \text{ because } \sigma_T^2 \text{ is a constant;}$$

$$\frac{\delta(\mathbf{b}'\mathbf{P}\mathbf{b})}{\delta \mathbf{b}} = 2\,\mathbf{P}\mathbf{b} \text{ and}$$

$$\frac{\delta(\mathbf{b}'\mathbf{c})}{\delta \mathbf{b}} = \mathbf{c}\,.$$

Thus

$$2\mathbf{P}\mathbf{b} - 2\mathbf{c} = 0 \text{ and } \mathbf{P}\mathbf{b} = \mathbf{c}$$

so that

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{c}\,.$$

For an animal with information vector, **x**, the index is:

$$I = \mathbf{b}'\mathbf{x}\,.$$

With the identity, $\mathbf{b} = \mathbf{P}^{-1}\mathbf{c}$,

σ_I^2 can be rewritten as :

$$\sigma_I^2 = \mathbf{b}'\mathbf{P}\mathbf{b} = \mathbf{c}'\mathbf{P}^{-1}\mathbf{P}\mathbf{P}^{-1}\mathbf{c} = \mathbf{c}'\mathbf{P}^{-1}\mathbf{c} = \mathbf{b}'\mathbf{c}.$$

The last expression in the series of equalities is the easiest to calculate.

Similarly, $\sigma_{TI} = \mathbf{b}'\mathbf{c}$ so that the correlation between I and T is:

$$r_{TI} = \sigma_{TI}/(\sigma_I^2\sigma_T^2)^{.5} = \mathbf{b}'\mathbf{c}/[(\mathbf{b}'\mathbf{c})(\sigma_T^2)]^{.5} = \sqrt{\mathbf{b}'\mathbf{c}/\sigma_T^2}.$$

If derivation of the selection index equations is from maximizing $r_{TI} = \sigma_{TI}/(\sigma_I^2\sigma_T^2)^{.5}$, the logarithm of r_{TI} is easier to work with. By remembering that

$$\frac{\delta[\log(x)]}{\delta(y)} = \left(\frac{1}{x}\right)\frac{\delta(x)}{\delta(y)}, \quad \text{then the parts of}$$

$$\frac{\delta[\log(\sigma_{TI}) - .5\ \log(\sigma_I^2) - .5\ \log(\sigma_T^2)]}{\delta \mathbf{b}} = 0$$

can be differentiated separately as follows:

$$\frac{\delta(\mathbf{b}'c)}{\delta b} = \mathbf{c}(1/\sigma_{TI})$$

$$\frac{\delta[.5\ \log(\mathbf{b}'\mathbf{P}\mathbf{b})]}{\delta \mathbf{b}} = \mathbf{P}\mathbf{b}(1/\sigma_I^2) \text{ and}$$

$$\frac{\delta[.5\ \log(\sigma_T^2)]}{\delta \mathbf{b}} = 0\ .$$

Thus

$$\mathbf{c}(1/\sigma_{TI}) - \mathbf{P}\mathbf{b}(1/\sigma_I^2) = 0 \text{ and } \mathbf{P}\mathbf{b} = \mathbf{c}(\sigma_I^2/\sigma_{TI})\ .$$

If the constant, σ_I^2/σ_{TI}, is set equal to 1, the equations are the same as for minimizing prediction error squared. A constant other than one could be chosen but would not change ranking by the index and would not change the r_{TI}. The calculations for σ_I^2 and σ_{TI} would, however, be different.

CHAPTER 8

DETERMINING THE COEFFICIENTS FOR SELECTION INDEX EQUATIONS

In matrix form, the selection index procedure is quite simple; an inversion and a matrix by vector multiply to solve for the weights, one vector by vector multiply for the index and another vector by vector multiply if accuracy or prediction error variance is wanted. These steps were shown in the appendix of Chapter 7. With or without matrix algebra, however, the difficult part of selection index procedures is to determine the numbers (coefficients in selection index jargon) that go into the left-hand sides (LHS) and right-hand sides (RHS) of the equations that must be solved to find the weights (**b**'s) for the records (X's). This chapter will utilize expected values to determine these coefficients from a few genetic parameters, such as heritability (h^2), repeatability (r), numbers of records for the different relatives, and the numerator relationships.

The X_i used in the selection index are often averages of records. The variance of an average depends partly on the covariance between records making up the average. Such covariances will be between records on the same animal or between records on relatives such as paternal half-sibs. An important step in finding the variance is to determine the covariance between records in the average.

MODELS FOR DETERMINING COVARIANCES BETWEEN RECORDS

Traits that can be measured only once can be represented by the model:

$$P_i = G_i + E_i$$

where P_i is the phenotypic record adjusted for fixed effects such as the overall mean,

G_i is the total genetic value, and

E_i is the total of all environmental effects.

The covariance between records on relatives i and j can be determined by expected values:

$$\text{Cov}(P_i,P_j) = E[(G_i + E_i)(G_j + E_j)] = \sigma_{G_iG_j} + \sigma_{E_iE_j} ,$$

under the usual assumption of no covariance between genetic and environmental effects. Note that $\sigma_{G_iG_j} = a_{ij}\sigma^2_{10} + d_{ij}\sigma^2_{01} + \cdots$ as developed in Chapter 6.

For convenience of notation, the covariance between environmental effects on records of relatives i and j will be defined as:

$$\sigma_{E_iE_j} = c_{ij}\sigma^2_X$$

where $\sigma^2_X = \sigma^2_P$ is the total or phenotypic variance. Thus, if only additive genetic effects are involved:

$$\sigma_{G_iG_j} = a_{ij}\sigma^2_{10} = a_{ij}h^2\sigma^2_X .$$

Then

$$\text{Cov}(P_i,P_j) = (a_{ij}h^2 + c_{ij})\sigma^2_X .$$

Even if other genetic effects are involved, this expression is often a good approximation for the phenotypic covariance.

Multiple measurement traits are those that allow repeated records, as for example a first milk lactation, a second milk lactation, etc. The model for such records is:

$$P_{ij} = G_i + PE_i + TE_{ij} ,$$

where P_{ij} is the jth phenotypic record of the ith animal adjusted for the mean and other fixed effects,

G_i is the total genetic value,

PE_i is the total of all permanent environmental effects which affect each record the animal makes, and

TE_{ij} is the total of all random temporary environmental effects which affect only the jth record of animal i.

This model may be an over-simplification of the true model for some multiple measurement traits but is often a reasonable approximation.

Because G_i and PE_i repeat in every record of the animal, this is sometimes called the repeated records or repeatability model and sometimes the animal model. The sum of all permanent effects of the animal can be denoted as the animal effect:

$$A_i = G_i + PE_i .$$

Repeatability, r, is defined as the fraction of the total variance which is due to animal effects:

$$r = \sigma_A^2/\sigma_X^2 = (\sigma_G^2 + \sigma_{PE}^2)/(\sigma_G^2 + \sigma_{PE}^2 + \sigma_{TE}^2) .$$

Note that $\sigma_A^2 = r\sigma_X^2$, an identity that is often useful.

The covariance between two records on the same animal is $\sigma_A^2 = r\sigma_X^2$ and can be determined with expected values:

$$\text{Cov}(P_{ij},P_{ij'}) = E[(A_i + TE_{ij})(A_i + TE_{ij'})] = \sigma_A^2 = r\sigma_X^2$$

under the assumption of zero covariances between animal effects and temporary environmental effects and between temporary environmental effects. Now the variance of an average which is the most important coefficient of the LHS's can be developed.

THE VARIANCE OF AN AVERAGE

Let X_i be the average of n_i records:

$$X_i = \frac{X_{i1} + \cdots + X_{in_i}}{n_i} .$$

If $E(X_{ij}^2) = \sigma_X^2$ for all i and j (that is, all records are from a distribution having the same variance) and if $E(X_{ij}X_{ij'}) = \sigma_{X'X}$ for all $j \neq j'$; that is, all pairs of records with a common i subscript have the same covariance, then:

$$\sigma_{X_i}^2 = E(X_i^2) = E\left(\frac{X_{i1} + \cdots + X_{in_i}}{n_i}\right)^2$$

$$= \frac{n_i\sigma_X^2 + n_i\,(n_i-1)\sigma_{X'X}}{n_i^2}$$

$$= \frac{\sigma_X^2 + (n_i-1)\sigma_{X'X}}{n_i} .$$

In the following paragraphs when X_i is the average of records on the same animal, $\sigma_{X'X}$ is the covariance between records on the same animal so that

$$\sigma_{X'X} = r\sigma_X^2.$$

When X_i is the average of single records on a group of equally related relatives (with additive and dominance relationships, $a_{ii'}$ and $d_{ii'}$), then $\sigma_{X'X}$ is the covariance between records of any pair of relatives i and i', each contributing a record to the average so that

$$\sigma_{X'X} = \sigma_{G_iG_{i'}} + c_{ii'}\sigma_X^2,$$

the sum of the total genetic covariance and the environmental covariance.

If $\sigma_{G_iG_{i'}} = a_{ii'}h^2\sigma_X^2$ (only additive genetic effects contribute to the genetic covariance), then:

$$\sigma_{X'X} = (a_{ii'}h^2 + c_{ii'})\sigma_X^2 .$$

COVARIANCE BETWEEN AVERAGES

The covariance between averages is often equal to the covariance between any record in the first average and any record in the other average. Expected values can be used to determine when this is true. Let X_{ik} be a record from average

$$X_i = \frac{X_{i1} + \cdots + X_{in_i}}{n_i},$$

and $X_{j\ell}$ be a record from average

$$X_j = \frac{X_{j1} + \cdots + X_{jn_j}}{n_j}.$$

If $E(X_{ik}X_{j\ell})$ is the same for all k and ℓ, then the expected value shows that in the numerator of

$$\text{Cov}(X_i, X_j) = E\left(\left(\frac{X_{i1} + \cdots + X_{in_i}}{n_i}\right)\left(\frac{X_{j1} + \cdots + X_{jn_j}}{n_j}\right)\right)$$

there are $n_i n_j$ expected values with the same expectation, $E(X_{ik}X_{j\ell})$, and that the denominator is $n_i n_j$. Thus, if a representative record from X_i is $G_i + E_i$ and a representative record from X_j is $G_j + E_j$, then

$$\begin{aligned}\text{Cov}(X_i, X_j) &= E[(G_i + E_i)(G_j + E_j)] \\ &= \sigma_{G_iG_j} + c_{ij}\sigma^2_X .\end{aligned}$$

SUMMARY OF VARIANCE OF AN AVERAGE

1) If X_i is the average of records on animal i, then $\sigma_{X'X} = r\sigma^2_X$,

2) If X_i is the average of single records of relatives of type i,

then $\sigma_{X'X} = \sigma_{G_iG_{i'}} + c_{ii'}\sigma^2_X$, and

3) If X_i is the average of n_i records on each of p_i relatives of type i,

then also $\sigma_{X'X} = \sigma_{G_iG_{i'}} + c_{ii'}\sigma^2_X$.

The derivation of the variance of an average of averages can be done with expected values using the often correct property that the covariance between averages is the same as the covariance between a record from one average and a record from the other average.

Let $X_i = \dfrac{\bar{X}_{i1} + \cdots + \bar{X}_{ip_i}}{p_i}$ where

$\bar{X}_{ij}$ is the average of n_i records on each animal j in relative group i. The number of animals in group i is p_i.

Then

$$\sigma^2_{X_i} = E\left(\frac{\bar{X}_{i1} + \cdots + \bar{X}_{ip_i}}{p_i}\right)^2 = \left(\frac{(\bar{X}_{i1})^2 + \cdots + (\bar{X}_{ip_i})^2 + \text{all products}}{p_i^2}\right)$$

$$= \frac{p_i V(\bar{X}_{ij}) + p_i(p_i-1)\text{Cov}(\bar{X}_{ij},\bar{X}_{ij'})}{p_i^2}$$

$$= \frac{\frac{\sigma^2_X + (n_i-1)r\sigma^2_X}{n_i} + (p_i-1)(\sigma_{G_iG_{i'}} + c_{ii'}\sigma^2_X)}{p_i}$$

4) This formula allows calculation of the variance of the average of any number of animals, each with any number of records using just a few parameters, i.e., variances do not need to be estimated for all combinations of number of animals and number of records from sets of data if the assumptions are correct for the previous derivation.

THE RIGHT HAND SIDES

The selection index weights (**b**'s) depend primarily on the variances of the averages and RHS's of the equations to find the selection index weights. If $\sigma^2_{X_i}$, $\sigma_{X_iX_j}$, and σ_{X_iT} are known for all i and j, the equations to find the appropriate weights for the index can be set up easily; $\sigma^2_{X_i}$ and $\sigma_{X_iX_j}$ can be estimated or derived as shown and do not depend on what is being predicted. The RHS's, σ_{X_iT}, however, are the covariances between what can be measured, the X_i, and T, something that cannot be measured or seen. Therefore, σ_{X_iT} must be computed indirectly. If selection is for additive genetic value, $\sigma_{X_iT} = a_{i\alpha}\sigma^2_{10}$ where $a_{i\alpha}$ is the additive relationship between the relative with record

X_i and the additive genetic value, G_{A_α}, of the individual α that is to be evaluated. The additive genetic variance is σ^2_{10}. The RHS is the portion of the genetic covariance between relatives **i** and **α** that is due to additive genetic effects in common. Recall that $\sigma^2_{10}/\sigma^2_X = h^2$, heritability in the "narrow sense". Thus, $\sigma^2_{10} = h^2\sigma^2_X$ and $\sigma_{X_iT} = a_{i\alpha}h^2\sigma^2_X$.

Although the usual case is to select for additive genetic value the selection index is more general and can be used for most possible definitions of T, the true value to be predicted. The only part of the selection index equations to find the weights that changes when T is redefined are the right-hand sides, the σ_{X_iT}. Of course, σ^2_T changes and other parameters that depend on σ^2_T and the RHS's will also change. Expected values and simple models can be used to find σ_{X_iT} and σ^2_T. The expected values will be demonstrated for several definitions of T, including the usual one where T is additive genetic value. To simplify the expected values, all variables will be assumed to have zero means, although, as stated earlier, variances and covariances are not affected by the means.

Case 1. $T = G_{A_\alpha}$, additive genetic value for animal α.

Let X_i be a representative record included in X_i with model $X_i = G_i + E_i$ or $X_i = G_i + PE_i + TE_i$, where G_i can also be separated into additive, dominance, additive by additive genetic values, etc; PE_i is the permanent environmental effect on all records of animal i; and TE_i is a temporary environmental effect on a specific record of i.

Then,
$$\sigma_{X_iT} = E(X_iG_{A_\alpha}) = E[(G_i + E_i)(G_{A_\alpha})]$$
$$= E(G_iG_{A_\alpha}) + E(E_iG_{A_\alpha})$$
$$= a_{i\alpha}\sigma^2_{10} + 0,$$

unless a nonzero genetic by environmental covariance exists. Thus, the right-hand sides of the selection index equations will be

$a_{i\alpha}\sigma^2_{10} = a_{i\alpha}h^2\sigma^2_X$, where σ^2_X is the phenotypic variance of individual records. If X_i is a record on animal α $(i=\alpha)$, then $a_{i\alpha} = 1 + F_\alpha$.

Similarly, $\sigma^2_T = E(G^2_{A_\alpha}) = a_{\alpha\alpha}\sigma^2_{10} = (1+F_\alpha)h^2\sigma^2_X$ if α is inbred and

$\sigma^2_T = \sigma^2_{10} = h^2\sigma^2_X$ if α is not inbred.

Case 2. $T = A_\alpha = G_\alpha + PE_\alpha$, real producing ability of animal α.

If $i = \alpha$, $\sigma_{X_iT} = E[(G_\alpha + PE_\alpha + TE_\alpha)(G_\alpha + PE_\alpha)]$

$= \sigma^2_G + \sigma^2_{PE} = \sigma^2_A = r\sigma^2_X$, if not inbred.

If $i \neq \alpha$, $\sigma_{X_iT} = E(G_i + PE_i + TE_i)(G_\alpha + PE_\alpha)]$

$= E(G_iG_\alpha + E(PE_iPE_\alpha)$

$= E(G_iG_\alpha) + E(PE_iPE_\alpha)$ + others likely to be zero

$= \sigma_{G_iG_\alpha} + \sigma_{PE_iPE_\alpha}$ which is the

total genetic covariance plus permanent environmental covariance that sometimes is assumed to be zero but is not necessarily so, e.g., for littermates. For all $i \neq \alpha$, and $\sigma^2_G = \sigma^2_{10}$, the RHS's will be the same as for predicting additive genetic value and if α has no records, the index weights and index will be the same as for predicting additive genetic value. However, σ^2_T will be different;

$\sigma^2_T = E(A^2_\alpha) = E[(G_\alpha + PE_\alpha)^2] = \sigma^2_A = \sigma^2_G + \sigma^2_{PE} = r\sigma^2_X$ if not inbred. Recall that repeatability or the correlation between records on

the same animal is defined as :

$$r = (\sigma_G^2 + \sigma_{PE}^2)/\sigma_X^2 = \sigma_A^2/\sigma_X^2.$$

Case 3. $T = G_{D_\alpha}$, dominance genetic value.

$$\sigma_{X_iT} = E[(G_i + E_i)(G_{D_\alpha})] = d_{i\alpha}\sigma_{01}^2 \ ;$$

$$\sigma_T^2 = E(G_{D_\alpha}^2) = \sigma_{01}^2 \ .$$

Case 4. $T = G_{A_\alpha} + G_{D_\alpha}$, additive plus dominance genetic value.

$$\sigma_{X_iT} = E[(G_i + E_i)(G_{A_\alpha} + G_{D_\alpha})] = a_{i\alpha}\sigma_{10}^2 + d_{i\alpha}\sigma_{01}^2 \ ;$$

$$\sigma_T^2 = E[(G_{A_\alpha} + G_{D_\alpha})^2] = \sigma_{10}^2 + \sigma_{01}^2 \text{ , if not inbred.}$$

Case 5. $T = G_\alpha$, overall genetic value.

$$\sigma_{X_iT} = E[(G_i + E_i)(G_\alpha)] = \sigma_{G_iG_\alpha} = a_{i\alpha}\sigma_{10}^2 + d_{i\alpha}\sigma_{01}^2 + \cdots \ ;$$

$$\sigma_T^2 = E(G_\alpha^2) = \sigma_G^2 = \sigma_{10}^2 + \sigma_{01}^2 + \cdots \text{ , if not inbred.}$$

If $G_\alpha = G_{A_\alpha}$, then case 5 is the same as predicting additive genetic value as in case 1.

Case 6. $T = (1/2)G_{A_\alpha}$, the average part of additive genetic value that is transmitted to progeny--transmitting ability. Transmitting ability is usually reported by most national dairy and beef sire and cow evaluations under such names as expected progeny difference (EPD), predicted difference (PD), and predicted transmitting ability (PTA or ETA).

$$\sigma_{X_iT} = E[(G_i + E_i)(1/2)(G_{A_\alpha})] = (1/2)E[(G_i + E_i)(G_{A_\alpha})]$$
$$= (1/2)a_{i\alpha}\sigma^2_{10}$$

Thus, because all RHS's are one-half those for predicting additive genetic value, the index weights and index will only be one-half as large as when predicting additive genetic value. The variance of T is:

$$\sigma^2_T = E[(1/2)^2(G^2_{A_\alpha})] = (1/4)E(G^2_{A_\alpha}) = (1/4)a_{\alpha\alpha}\sigma^2_{10}$$
$$= (1/4)\sigma^2_{10} \text{ for } F_\alpha = 0.$$

Thus, the r_{TI} will be the same as for predicting additive genetic value. These first six definitions show the flexibility of the selection index if T can be defined. In the following cases, there is more difficulty in determining exactly what T is.

Case 7. $T = P_\alpha = G_\alpha + E_\alpha = G_\alpha + PE_\alpha + TE_\alpha$, a future record (this is probably what most breeders think is happening in cases, 1, 2, and 5).

If $i = \alpha$ (animal already has a record, e.g., record $P_{\alpha 1}$ and want to predict from this record, record $P_{\alpha 2}$),

$$\sigma_{X_iT} = E[(G_\alpha + PE_\alpha + TE_{\alpha 1})(G_\alpha + PE_\alpha + TE_{\alpha 2})]$$
$$= \sigma^2_G + \sigma^2_{PE} = r\sigma^2_X,$$

if not inbred.

If $i \neq \alpha$,

$$\sigma_{X_iT} = E[(G_i + PE_i + TE_i)(G_\alpha + PE_\alpha + TE_\alpha)]$$
$$= E(G_iG_\alpha) + E(PE_iPE_\alpha) + E(TE_iTE_\alpha)$$
$$= Cov(G_iG_\alpha) + Cov(E_iE_\alpha).$$

The two environmental covariances sometimes can be assumed to be zero. The first term is the total genetic covariance and not just the covariance due to additive genetic effects. These right-hand sides and index weights are the same as for predicting real producing ability if $E(TE_i TE_\alpha) = 0$. However, σ^2_T is different;

$$\sigma^2_T = E(G_\alpha + PE_\alpha + TE_\alpha)^2 = \sigma^2_G + \sigma^2_{PE} + \sigma^2_{TE} = \sigma^2_G + \sigma^2_E = \sigma^2_X,$$

the total or phenotypic variance of single records.

Case 8. T = average of records of **m** future half-sib progeny of some sire

$$= [(\Sigma G_i)/m] + [(\Sigma E_i)/m]$$

Because the covariance between averages and between individual records is the same in this case, let

$P_\alpha = G_\alpha + E_\alpha$ be a representative record in T; then

$$\sigma_{X_i T} = E[(G_i + E_i)(G_\alpha + E_\alpha)] = \sigma_{G_i G_\alpha}$$
$$(= a_{i\alpha}\sigma^2_{10} \text{ if } G_\alpha = G_{A_\alpha}).$$

However,

$$\sigma^2_T = E\{[\frac{\Sigma(G_i + E_i)}{m}]^2\} = \frac{\sigma^2_X + (m-1)\sigma_{X'X}}{m},$$

where $\sigma_{X'X}$ is the covariance between pairs of records in the average T. This term can be evaluated as before and will have one or more genetic components and possibly an environmental covariance,

$$\sigma_{G_i G_{i'}} + \sigma_{E_i E_{i'}},$$

in which the genetic plus environmental covariance between i and i' are both included.

Case 9. T = average of records of an infinite number, ∞, of future half-sib progeny of some sire.

σ_{X_iT} is as in (case 8), but

$\sigma_T^2 = (\sigma_X^2/m) + [(m-1)\sigma_{X'X}/m]$ and as $m \to \infty$, $\sigma_T^2 \to \sigma_{X'X}$

where

$$\sigma_{X'X} = \sigma_{G_iG_{i'}} + \sigma_{E_iE_{i'}}.$$

When $m = \infty$, this case is the same as predicting $(1/2)(G_{A_\alpha})$ of a sire if $G_\alpha = G_{A_\alpha}$.

Case 10. T = average additive genetic value of **m** or ∞ future half-sib progeny, $(\Sigma G_{A_{\alpha j}})/m$.

$\sigma_{X_iT} = a_{i\alpha}\sigma_{10}^2$ as in (case 8), and

$\sigma_T^2 = E\{[(\Sigma G_{A_\alpha})/m]^2\} = [\sigma_{10}^2 + (m-1)a_{\alpha\alpha'}\sigma_{10}^2]/m$

because σ_{10}^2 is the variance of additive genetic values and $a_{\alpha\alpha'}\sigma_{10}^2$ is the covariance between additive genetic values of α and α', a representative pair in the group. As $\mathbf{m} \to \infty$, $\sigma_T^2 \to a_{\alpha\alpha'}\sigma_{10}^2$. For noninbred half-sib progeny, $a_{\alpha\alpha'} = 1/4$ and $\sigma_T^2 = (1/4)\sigma_{10}^2$ as in case 6 when predicting $(1/2)(G_{A_{\text{sire}}})$. In case 10, α refers to a progeny group, and in case 6, α designates a particular sire that has the progeny.

These examples illustrate the power of the selection index method; T can be almost anything, even, for example, difference in additive genetic value between animals or linear functions of genetic values. The absolute necessity of clearly defining what T is should be clear. Precise definition of T would avoid much confusion.

AVERAGE OF RECORDS OF A SINGLE RELATIVE

If X_i is the average of n_i records on an animal, then the variance of the average can be found as a function of variances and covariances of the records going into the average. If, as often is nearly true, the variance of first records equals the variance of second records, etc., and the covariances are all equal, then

$$\sigma^2_{X_i} = \sigma^2_X \left(\frac{1 + (n_i - 1)r}{n_i} \right),$$

where σ^2_X is the variance associated with single records and **r** is repeatability. Thus, the diagonal coefficient of the selection index equations to find the selection index weights is:

$$\sigma^2_X \left(\frac{1 + (n_i - 1)r}{n_i} \right).$$

Each off-diagonal coefficient is the same as the covariance between a single record of one animal and a single record of another relative.

If, however, the only reason for likeness between relatives is common additive genetic effects, then the off-diagonal coefficients are of the form $\sigma_{X_i X_j} = a_{ij}\sigma^2_{10} = a_{ij}h^2\sigma^2_X$. If other components of genetic variance are important, this expression is not the true covariance but may be a reasonable approximation because the coefficients of the other components will be small. A more likely source of error is the possibility of an environmental covariance among relatives. If $c_{ij}\sigma^2_X$ is the covariance between records of relatives **i** and **j** caused by common environmental effects, then the off-diagonal coefficients should be $\sigma_{X_i X_j} = (a_{ij}h^2 + c_{ij})\sigma^2_X$. The equations to find the **b**'s can be written (assuming all $c_{ij} = 0$) to predict G_{A_α}:

$$\left(\frac{1+(n_1-1)r}{n_1}\right)\sigma_X^2 b_1 + a_{12}h^2\sigma_X^2 b_2 + \cdots + a_{1N}h^2\sigma_X^2 b_N = a_{1\alpha}h^2\sigma_X^2$$

$$a_{12}h^2\sigma_X^2 b_1 + \left(\frac{1+(n_2-1)r}{n_2}\right)\sigma_X^2 b_2 + \cdots + a_{2N}h^2\sigma_X^2 b_N = a_{2\alpha}h^2\sigma_X^2$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$a_{1N}h^2\sigma_X^2 b_1 + a_{2N}h^2\sigma_X^2 b_2 + \cdots + \left(\frac{1+(n_N-1)r}{n_N}\right)\sigma_X^2 b_N = a_{N\alpha}h^2\sigma_X^2$$

Because σ_X^2 appears in each equation, dividing each equation by σ_X^2 will not change the solutions for the **b**'s.

Thus, the equations can be written as:

$$\frac{1+(n_1-1)r}{n_1}b_1 + a_{12}h^2 b_2 + \cdots + a_{1N}h^2 b_N = a_{1\alpha}h^2$$

$$a_{12}h^2 b_1 + \frac{1+(n_2-1)r}{n_2}b_2 + \cdots + a_{2N}h^2 b_N = a_{2\alpha}h^2$$

$$\vdots \qquad\qquad\qquad\qquad\qquad \vdots$$

$$a_{1N}h^2 b_1 + a_{2N}h^2 b_2 + \cdots + \frac{1+(n_N-1)r}{n_N}b_N = a_{N\alpha}h^2$$

Only **r** and $\mathbf{h}^2$ are necessary in order to set up the equations because the relationships can be computed and the **n**'s will be known. These equations are sometimes called the simplified equations . Another simplification is to divide by $h^2\sigma_X^2$.

AVERAGE OF n RECORDS ON p EQUALLY RELATED RELATIVES

Let X_i be the average of a genetic group of animals ($\mathbf{p}_i$) each with $\mathbf{n}_i$ records (e.g., a group of paternal half-sisters each with 2 records). Further, 1) each animal in the group has the same relationship, $a_{ii'}$, to all other animals in the group, and 2) each animal in group **i** has the same relationship to all animals in group **j**, i.e., a_{ij} is same for all pairs of animals; one from group **i** and one from group **j**. Then the diagonal coefficients become:

$$\sigma^2_{X_i} = \left(\frac{\frac{1+(n_i-1)r}{n_i} + (p_i - 1)a_{ii'}h^2}{p_i} \right) \sigma^2_X .$$

If other than additive genetic variance contributes to likeness between animals in the genetic group, the part of the numerator corresponding to the covariance within group will be greater. For example, if the environmental covariance is $c_{ii'}\sigma^2_X$ and there is also likeness due to dominance genetic variance, the diagonal coefficients are:

$$\sigma^2_{X_i} = \left(\frac{\frac{1+(n_i-1)r}{n_i} + (p_i-1)(a_{ii'}\, h^2 + d_{ii'}\sigma^2_{01}/\sigma^2_X + c_{ii'})}{p_i} \right) \sigma^2_X .$$

After dividing by σ_X^2 and with the assumptions to be again stated, the simplified equations for finding the appropriate weights for the index

$$\hat{G}_{A_\alpha} = I = \sum_{i=1}^{N} b_i X_i \qquad \text{are:}$$

$$d_1 b_1 + a_{12}h^2 b_2 + \cdots + a_{1N}h^2 b_N = a_{1\alpha}h^2$$

$$a_{12}h^2 b_1 + d_2 b_2 + \cdots + a_{2N}h^2 b_N = a_{2\alpha}h^2$$

$$\vdots \qquad\qquad\qquad\qquad\qquad\qquad \vdots$$

$$a_{1N}h^2 b_1 + a_{2N}h^2 b_2 + \cdots + d_N b_N = a_{N\alpha}h^2$$

where $d_i = \dfrac{\dfrac{1+(n_i-1)r}{n_i} + (p_i - 1)a_{ii'}h^2}{p_i}$.

The assumptions that are implied by this simplified set of equations are:

1) selection is for additive genetic value,
2) the variances of single records for all relatives are σ_X^2,
3) the covariance between records on an animal is $r\sigma_X^2$ for all relatives.
4) only additive genetic variance contributes to the covariance among relatives. If this assumption is not true, the $a_{ij}h^2$ terms should be modified to take into account other components of genetic variance and any environmental covariance, and
5) each animal in group i has the same number of records. If not, the group should be divided so that each animal in a sub-group has the same number of records.

RECORDS FROM INBRED ANIMALS

The variation among inbred animals will be greater than for non-inbred animals since the genetic variance of inbred animals is $a_{ii}\sigma^2_{10} = (1+F_i)\sigma^2_{10}$ when only additive genetic effects are considered. Thus, the phenotypic variance among single records of inbred animals is $(1+F_i)\sigma^2_{10} + \sigma^2_E = [(1+F_i)h^2 + (1-h^2)]\sigma^2_X$, where σ^2_X is the variance of single records of non-inbred animals. The diagonal coefficients of the equations that determine the selection index weights will be increased. For single records, the increase will be $F_ih^2\sigma^2_X$ to $(1+F_ih^2)\sigma^2_X$. For the average of records on the same animal, the diagonal coefficient will be:

$$\left(\frac{1+(n_i-1)r}{n_i} + F_ih^2\right)\sigma^2_X,$$

because the covariance between records on the animal will also increase by $F_ih^2\sigma^2_X$. For the average of single records on each of p_i animals in group **i**, the diagonal coefficient will be:

$$\left(\frac{(1+F_ih^2) + (p_i-1)a_{ii'}h^2}{p_i}\right)\sigma^2_X .$$

Although that situation seems rather unlikely, in most such cases, $a_{ii'}$ will be larger than if the animals were not inbred. For the average of $\mathbf{n_i}$ records on each of $\mathbf{p_i}$ animals in group **i**, the diagonal coefficient becomes:

$$\left\{\frac{\left(\frac{1+(n_i-1)r}{n_i} + F_ih^2\right) + (p_i-1)a_{ii'}h^2}{p_i}\right\}\sigma^2_X .$$

If animal α is inbred, the formula for the denominator of r_{TI} will be $\sigma^2_T = (1+F_\alpha)h^2\sigma^2_X$ when selecting for G_{A_α}

COMPUTATION OF ACCURACY WITH SIMPLIFIED EQUATIONS

The solutions for the **b**'s will be the same for the simplified equations as for the regular equations if the assumptions are true. For the regular equations

$$r_{TI}^2 = \Sigma b_i \sigma_{X_i T} / \sigma_T^2 .$$

For the simplified equations, substitute $a_{i\alpha} h^2 \sigma_X^2 = a_{i\alpha} \sigma_{10}^2$ for $\sigma_{X_i T}$ and remember that $\sigma_T^2 = \sigma_{10}^2$ if T is additive genetic value.

Then

$$r_{TI}^2 = \Sigma b_i a_{i\alpha} \sigma_{10}^2 / \sigma_{10}^2 = \Sigma b_i a_{i\alpha} \text{ and } r_{TI} = \sqrt{\Sigma b_i a_{i\alpha}}$$

Thus, only **r** and $\mathbf{h}^2$ are needed to compute the **b**'s and $\mathbf{r}_{TI}$ with the simplified equations when selecting for additive genetic value.

VARIANCE OF T GIVEN THE INDEX WITH SIMPLIFIED EQUATIONS

$\sigma_{10}^2 = \sigma_T^2$ will be needed since

$$\sigma_{T|I=I_0}^2 = (1-r_{TI}^2)\sigma_T^2 = (1-\Sigma b_i a_{i\alpha})\sigma_T^2 = (1-\Sigma b_i a_{i\alpha}) h^2 \sigma_X^2 .$$

ADDITIONAL NOTE

Often all animals will not have records available on the same types of relatives. Even when records are available on the same relatives, the relatives may not have the same number of records. The selection index procedure can still be used to compare animals, but then the weights for the index for each animal with a different set of records and types of relatives will have to be found from the set of equations corresponding to the $\mathbf{p}_i$'s and $\mathbf{n}_i$'s associated with records of relatives of that animal.

APPLICATION OF THE INDEX TO CASES WHERE THE ASSUMPTIONS ARE TRUE

1. One or several records on the individual being evaluated.

Often individuals must be compared on the basis of their performance but with unequal numbers of records. The best procedure is to solve the index equations for each specific case (i.e., number of records per individual). If, however, all the variances = σ_X^2, all the covariances among the X's = $r\sigma_X^2$, and the covariances on the RHS's all are equal, then the equations can be simplified.

If the covariances between all records and the additive genetic value of the individual are all equal, as is a common assumption, the index becomes

$$I = \mathbf{b}X$$

where X is the average of **n** records on the individual to be indexed for additive genetic value.

The equation to find **b** for equal variances and covariances is:

$$\left(\frac{1 + (n-1)r}{n}\right) b = h^2 \quad \text{so that } b = \frac{nh^2}{1+(n-1)r}.$$

$$r_{TI} = \sqrt{\left(\frac{nh^2}{1+(n-1)r}\right)\left(h^2\sigma_X^2/h^2\sigma_X^2\right)} = \sqrt{\frac{nh^2}{1+(n-1)r}} \quad \text{because} \quad \sigma_T^2 = \sigma_{10}^2 = h^2\sigma_X^2.$$

Then

$$\sigma^2_{T|I=I_0} = \left(1 - \frac{nh^2}{1+(n-1)r}\right) h^2\sigma_X^2$$ for animals with the same number of records and the same index value, I_0.

This procedure allows animals with varying numbers of records to be ranked according to estimated breeding value so that the probability of correctly ranking the animals is maximized.

2. The case of using one record on each of many relatives to estimate the breeding value of animal α.

The index equations will be:

$$\begin{aligned}
b_1 + a_{12}h^2b_2 + \cdots + a_{1N}h^2b_N &= a_{1\alpha}h^2 \\
a_{12}h^2b_1 + \quad b_2 + \cdots + a_{2N}h^2b_N &= a_{2\alpha}h^2 \\
\vdots \qquad\qquad\qquad\qquad\qquad &\quad \vdots \\
a_{1N}h^2b_1 + a_{2N}h^2b_2 + \cdots + \quad b_N &= a_{N\alpha}h^2 .
\end{aligned}$$

Only additive relationships and heritability are needed to set up the equations to solve for the selection index weights.

3. The case where related individual **i** has more than one record ($\mathbf{n}_i$).

Now the diagonal coefficient will be:

$$\frac{1+(n_i-1)r}{n_i} .$$

The off-diagonals and RHS's will be the same as case 2.

4. The case where the X_i are the averages of single records of $\mathbf{p}_i$ members of group **i** with relationship $a_{ii'}$ with each other and all having the same relationship to α and to other groups or individuals used in the index.

Now the diagonal coefficients will be:

$$\frac{1+(p_i-1)a_{ii'}h^2}{p_i} .$$

The off-diagonals and RHS's will be the same as in cases 2 and 3.

5. The case where $\mathbf{p}_i$ members of group **i** have more than one record (n_i).

The diagonal coefficients will be:

$$\frac{\frac{1+(n_i-1)r}{n_i} + (p_i-1)a_{ii'}h^2}{p_i}.$$

The off-diagonal coefficients and RHS's are the same as before. This case provides the general form of the diagonal coefficients because when $\mathbf{n}_i = 1$, the diagonal is the same as for case 4; when $\mathbf{p}_i = 1$, the diagonal is the same as for case 3 and when $\mathbf{n}_i = 1$ and $\mathbf{p}_i = 1$, the diagonal coefficient is the same as case 2.

6. If members of a group of related individuals have differing numbers of records, then each subgroup with different numbers of records per individual can be treated as a separate group.

APPROXIMATION TO THE SELECTION INDEX WHEN h^2 IS SMALL

If heritability is small, a further approximation can be made to the selection index equations. The simplified equations have $a_{ij}h^2$ as the off-diagonal coefficients. The a_{ij}'s are less than or equal to 1/2 except for unusual situations. If $h^2 = .05$, then all the off-diagonal coefficients are less than or equal to (1/2) (.05) = (1/40). The approximation is to set these small off-diagonal coefficients to zero. The equations then become:

$$\begin{aligned} d_1b_1 &= a_{1\alpha}h^2 \\ d_2b_2 &= a_{2\alpha}h^2 \\ \cdot & \quad \cdot \\ \cdot & \quad \cdot \\ \cdot & \quad \cdot \\ d_Nb_N &= a_{N\alpha}h^2 \end{aligned}$$

where $d_1, d_2, \cdots, d_N$ are the diagonal coefficients after dividing by σ_X^2.

With this approximation, the weights are proportional to the relationships when each relative has only one record. This procedure also provides an approximate computational check for cases where the off-diagonals are really greater than zero. The r_{TI} appears larger than it really is if the r^2_{TI} is computed as $\Sigma b_i a_{i\alpha}$ since the **b**'s will be larger than they should be.

The true r_{TI} will be $\sigma_{TI}/(\sigma_I^2\sigma_T^2)^{.5}$ where $\sigma_{TI} = E(TI)$ and $\sigma_I^2 = E(I^2)$.

SELECTION INDEX NOT EXPRESSED IN DEVIATIONS

So far the index has been expressed in deviations where I as well as the X's are deviations from their population averages.

This is equivalent to

$$I_{deviation} = (I - \mu_o) = b_1(X_1-\mu_1) + b_2(X_2-\mu_2) + \cdots + b_N(X_N-\mu_N).$$

If the index is desired as an actual value, then if the μ's are known,

$$I = \mu_o + b_1(X_1-\mu_1) + \cdots + b_N(X_N-\mu_N)$$

where I is not expressed as a deviation; μ_o is the average of the population where the animals being indexed will make records; and μ_i, for $i=1, \cdots, N$, are the population averages associated with the records of the various relatives used in the index.

Example: Suppose a dairy cow makes a record of 14000 in a herd that averages 12000. A progeny record in that herd is to be predicted.

Then if $h^2 = .25$

$$b = (1/2)h^2 = .125$$

$$I_{progeny} = 12000 + .125(14000 - 12000) = 12{,}250 .$$

Suppose instead that the herd average will increase to 13000 before the progeny makes a record. Then

$$I_{progeny} = 13000 + .125(14000 - 12000) = 13{,}250 .$$

TABLE 8.1. WEIGHTS AND ACCURACY VALUES FOR PREDICTING ADDITIVE GENETIC VALUE FROM RECORDS OF VARIOUS RELATIVES. (h^2 IS HERITABILITY; r IS REPEATABILITY).

Records		Selection Index Weights	Accuracy = r_{TI}
Individual	(1)	h^2	$\sqrt{h^2}$
	(n)	$nh^2/[1 + (n-1)r]$	$\sqrt{nh^2/[1 + (n-1)r]}$
Dam or sire or progeny	(1)	$h^2/2$	$.50\sqrt{h^2}$
	(n)	$nh^2/[1 + (n-1)r](2)$	$.50\sqrt{nh^2/[1 + (n-1)r]}$
Sire and dam	(1)	$h^2/2$; $h^2/2$	$.71\sqrt{h^2}$
	(n)	$.5nh^2/[1 + (n-1)r]$; $.5nh^2/[1 + (n-1)r]$	$.71\sqrt{nh^2/[1 + (n-1)r]}$
One grandparent		$h^2/4$	$.25\sqrt{h^2}$
Four grandparents		All $h^2/4$	$.50\sqrt{h^2}$
One great-grandparent		$h^2/8$	$.125\sqrt{h^2}$
Eight great-grandparents		All $h^2/8$	$.35\sqrt{h^2}$
Individual and one parent or progeny		$[h^2-(h^2/2)^2]/[1 - (h^2/2)^2]$; $[h^2(1-h^2)/2]/[1 - (h^2/2)^2]$	$\sqrt{(5h^2-2h^4)/(4-h^4)}$
Individual and both parents		$h^2(h^2-2)/(h^4-2)$; $h^2(h^2-1)/(h^4-2)$ •••	$\sqrt{h^2(2h^2-3)/(h^4-2)}$
Individual and one grandparent or grandprogeny		$h^2(h^2-16)/(h^4-16)$; $4h^2(h^2-1)/(h^4-16)$	$\sqrt{h^2(2h^2-17)/(h^4-16)}$
Individual and four grandparents		$h^2(h^2-4)/(h^4-4)$; $h^2(h^2-1)/(h^4-4)$ •••	$\sqrt{h^2(2h^2-5)/(h^4-4)}$
Parent and progeny		$2h^2/(4+h^2)$; $2h^2/(4+h^2)$	$\sqrt{2h^2/(4+h^2)}$
Progeny (p half-sibs)		$2ph^2/[4 + (p-1)h^2]$	$\sqrt{ph^2/[4 + (p-1)h^2]}$

TABLE 8.1 continued:

Let $A = [1 + (n-1)r]/n$, $D = \{1 + [(p-1)h^2/4]\}/p$, and $C = AD - (h^4/16)$.

Records	Weights	Accuracy
Individual (n) and paternal half-sibs (p)	$[h^2D - (h^2/4)^2]/C$; $h^2(A-h^2)/4C$	$\sqrt{b_1 + (b_2/4)}$
Individual (n) and his paternal half-sib progeny (p)	$[h^2D - (h^2/2)^2]/[C - (3h^4/16)]$; $.5h^2(A-h^2)/[C - (3h^4/16)]$	$\sqrt{b_1 + (b_2/2)}$
Dam (n) and paternal half-sibs (p)	$.5nh^2/[1 + (n-1)r]$; $ph^2/[4 + (p-1)h^2]$	$\sqrt{b_1/2 + (b_1/4)}$
Dam (1) sire (1), and progeny (1)	$[h^2 - (h^4/16)]/[2 - (h^4/64)]$; $[h^2 - (h^4/16)]/[2 - (h^4/64)]$; $[h^2 - (h^4/8)]/[2 - (h^4/64)]$	$\sqrt{(b_1 + b_2 + b_3)/2}$
Paternal half-sibs (m), dam (n), and dam's paternal half-sibs (p)	$mh^2/[4 + (m-1)h^2]$; $h^2[D - (h^2/16)]/(2C)$ $h^2(A-h^2)/(8C)$	$\sqrt{b_1/4 + b_2/2 + b_3/8}$

CHAPTER 9

SIRE EVALUATION, EXAMPLE OF APPLICATION OF SELECTION INDEX

Many traits cannot be measured on males, thus genetic evaluation must be based either on records of female ancestors or on records of female progeny. Evaluation on the basis of progeny also usually results in much greater accuracy (r_{TI}) than pedigree evaluation, even with traits measured on both sexes. This method has received much use in dairy cattle and poultry breeding and also with other classes of animals.

ESTIMATION OF BREEDING VALUE

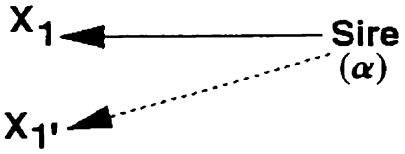
The problem of prediction of breeding value from progeny records in the simplest form is that the average of single records adjusted for fixed factors of **p** progeny all from different dams, X_1, is known and the additive genetic value of sire, α, is to be predicted as:

$$I = b_1 X_1.$$

If the assumptions discussed earlier are true, the simplified equation to find the best weighting factor, b_1, is:

$$\left[\frac{1 + (p-1)\, a_{11'} h^2}{p}\right] b_1 = a_{1\alpha} h^2.$$

In the diagram



X_1 and $X_{1'}$ are a representative pair of records in the progeny average and α represents the animal to be evaluated.

In this situation, $a_{11'} = .25$ and $a_{1\alpha} = .5$, so that:

$$b_1 = \frac{.5\,ph^2}{1+(p-1)(.25\,h^2)} = \frac{2p}{p+\frac{4-h^2}{h^2}} = \frac{2p}{p+\lambda} \text{ for } \lambda = (4-h^2)/h^2.$$

As $p \longrightarrow \infty$, $b_1 \longrightarrow 2.0$.

For $h^2 = .25$, $b_1 = \frac{2p}{p+15}$; for $h^2 = .5$, $b_1 = \frac{2p}{p+7}$; etc.

In general, $r_{TI} = \sqrt{a_{1\alpha}b_1} = \sqrt{\frac{p}{(p+\lambda)}}$. Note that

as the number of progeny, $p \longrightarrow \infty$, $r_{TI} \longrightarrow 1.00$.

Note:

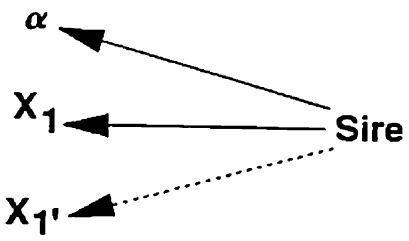
1) a new equation is not needed for each sire with a different number of progeny, because $\mathbf{b}_1$ has been solved for in terms of **p** and h^2,
2) b_1 depends on **p**,
3) r_{TI} depends on **p**, and
4) b_1 can exceed one. For most genetic evaluations, the **b**'s are usually less than one except for sire evaluation from progeny records.

VARIATIONS ON SIRE EVALUATION

The preceding section describes the basis for predicting additive genetic value of a sire from his progeny. Similar procedures that have been used will yield a weighting factor which is one-half this b_1; for example, $\frac{p}{p+15}$ rather than $\frac{2p}{p+15}$.

The following two additional definitions of true value result in the smaller weight.

Definition II. Rather than estimating the breeding value of the sire, the breeding value of a future progeny, α, is to be estimated.



$a_{11'} = .25$ as before, but $a_{1\alpha} = .25$ rather than .5. Again, $\lambda = (4-h^2)/h^2$.

Now, $b_1 = \frac{p}{p+\lambda}$ and, in this case, as $p \longrightarrow \infty$, $b_1 \longrightarrow 1.0$.

For $h^2 = .25$, $b_1 = \frac{p}{p+15}$; for $h^2 = .5$, $b_1 = \frac{p}{p+7}$; etc.

Also $r_{TI} = \sqrt{.25\left(\frac{p}{p+\lambda}\right)} = .5\sqrt{\frac{p}{p+\lambda}}$

and, in this case, as $p \longrightarrow \infty$, $r_{TI} \longrightarrow .5$.

Note that this accuracy is for predicting the additive genetic value of an animal from records of **p** paternal half sibs.

Definition III. The daughter or progeny superiority of a sire (also called transmitting ability) is to be predicted. Progeny superiority is defined to be the average of an infinite number of future progeny which is equivalent to one-half the additive genetic value of the sire, i.e., $T = .5\, G_{SIRE}$.

Then, $\sigma_{X_1T} = .5\, a_{1\alpha} h^2 \sigma_X^2 = .25\, h^2 \sigma_X^2 = .25\, \sigma_{10}^2$ because $a_{1\alpha} = .5$.

The equation to find b_1 is:

$$\left[\frac{1 + (p-1)\ .25\ h^2}{p}\right] b_1 = (.5)(.5)(h^2) \quad \text{and } b_1 = \frac{p}{p+\lambda} \quad \text{as in definition II.}$$

The accuracy, r_{TI}, however, will not be the same as in definition II but will be the same as for predicting the breeding value of the sire.

Remember $r_{TI} = \sqrt{\dfrac{b_1 \sigma_{X_1T}}{\sigma_T^2}}$. Note that $b_1 \sigma_{X_1T} = \left(\dfrac{p}{p+\lambda}\right)\ (.5)\ (.5)\ \sigma_{10}^2$.

But $T = .5\, G_{Sire}$. Thus, because G is additive value:

$$\sigma_T^2 = E(T^2) = E[(G/2)^2] = (.25)E[\, G^2] = (.25)\sigma_{10}^2 .$$

Thus,

$$r_{TI} = \sqrt{\frac{[p/(p+\lambda)](.25\ \sigma_{10}^2)}{.25\sigma_{10}^2}} = \sqrt{\frac{p}{p+\lambda}}$$

which is the same r_{TI} as when estimating the additive genetic value of the sire. This result should be expected because the only change has been to divide what is to be predicted by a constant one-half. The only difference in the evaluations is a factor of one-half. Ranking will be the same. The important point is to define T exactly, since what T is, makes a difference in the weighting factor and may make a difference in r_{TI} .

ENVIRONMENTAL COVARIANCE IN SIRE EVALUATION

If progeny are treated more alike because they are related than are unrelated animals, then an environmental covariance in addition to a genetic covariance will exist among animals in a progeny group. Assume that the environmental correlation among half-sibs in the same environment is $c_{11'}\sigma_X^2$.

The equation to find b_1 to evaluate the sire from p progeny with one record each is:

$$\left[\frac{1 + (p-1)(a_{11'}h^2 + c_{11'})}{p}\right] b_1 = a_{1\alpha}h^2 \quad ,$$

where $a_{11'}$ is the relationship among animals in the group, ($a_{11'} = .25$, if half sibs),

$c_{11'}\sigma_X^2$ is the environmental covariance, and

$a_{1\alpha}$ is relationship of animals in the group to α. If α is the sire, then $a_{1\alpha} = .5$.

Thus, $b_1 = \dfrac{.5\ ph^2}{1 + (p-1)(.25\,h^2 + c_{11'})}$ and $r_{TI} = \sqrt{\dfrac{.25\,ph^2}{1 + (p-1)(.25\,h^2 + c_{11'})}}$.

If $c_{11'} = .25\ h^2$ and $h^2 = .25$ as is approximately true for lactation yield:

$b = \dfrac{p}{p+7}$ or $\dfrac{2p}{p+14}$ rather than $\dfrac{2p}{p+15}$ with no environmental correlation,

and $r_{TI} = \sqrt{\dfrac{.5\ p}{p + 7}} = .71\sqrt{\dfrac{p}{p + 7}}$.

In this case as $p \longrightarrow \infty$, $b_1 \longrightarrow 1$, but $r_{TI} \longrightarrow .71$.

The important point is if $c_{11'} \neq 0$, then as $p \longrightarrow \infty$, $r_{TI} \longrightarrow$ less than 1, depending on the ratio, $c_{11'}/a_{11'}h^2$.

The following table compares the b_1's and r_{TI}'s when $h^2 = .25$ with and without environmental correlation.

Number of progeny p	No environmental correlation $c_{11'} = 0$		Environmental correlation $c_{11'} = .0625$	
	$b_1 = \frac{2p}{p+15}$	$r_{TI} = \sqrt{\frac{p}{p+15}}$	$b_1 = \frac{p}{p+7}$	$r_{TI} = .71\sqrt{\frac{p}{p+7}}$
1	.125	.25	.125	.25
3	.33	.41	.30	.39
10	.80	.62	.59	.54
20	1.15	.76	.74	.61
50	1.54	.88	.88	.66
100	1.73	.93	.93	.69
1000	1.98	.99	.99	.70
∞	2.00	1.00	1.00	.71

The previous table assumes the environmental correlation is the same for all pairs of progeny. The USDA dairy sire evaluation procedures in the past and now with mixed model methods, however, assume only records of daughters of a sire in the same herd have an environmental correlation. If there are n_i daughters in the i^{th} herd, then:

$$b_1 = \frac{2\,ph^2}{4 + (p-1)\,h^2 + \frac{4\Sigma n_i\,(n_i-1)\,c_{11'}}{p}}$$. If $h^2 = .25$ and $c_{11'} = .0625$,

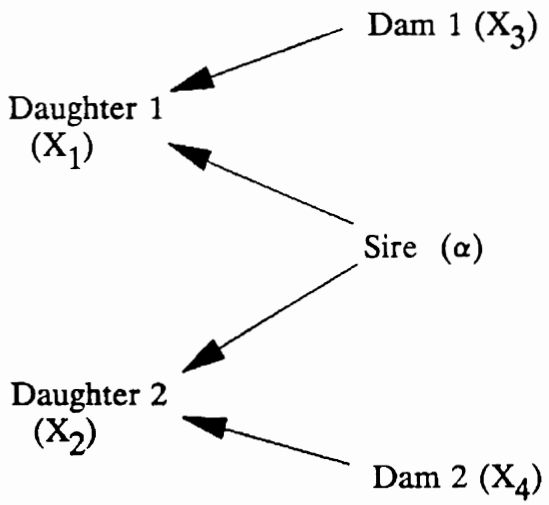
$$b_1 = \frac{2p}{p + 15 + \frac{\Sigma n_i\,(n_i-1)}{p}}$$ as compared to $b_1 = \frac{2p}{p+15}$ with $c_{11'} = 0$.

As before, $r_{TI} = \sqrt{.5\,b_1}$; and for $h^2 = .25$ and $c_{11'} = .0625$,

$$r_{TI} = \sqrt{\frac{p}{p+15 + \frac{\Sigma n_i\,(n_i-1)}{p}}}.$$

CORRECTION FOR LEVEL OF MATES

If mates of one sire are much superior to the mates of another sire, then this knowledge could be used in evaluating the sires from their progeny averages to avoid bias from the selected mates. One approach is to set up one equation for each daughter and one equation for each dam record. For two dams and two daughters:



The equations to find the b's for $I = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4$ are :

$$
\begin{aligned}
b_1 + .25\,h^2b_2 + .5\,h^2b_3 + 0\,b_4 &= .5\,h^2 \\
.25\,h^2b_1 + b_2 + 0\,b_3 + .5\,h^2b_4 &= .5\,h^2 \\
.5\,h^2b_1 + 0\,b_2 + b_3 + 0\,b_4 &= 0 \\
0\,b_1 + .5\,h^2b_2 + 0\,b_3 + b_4 &= 0
\end{aligned}
$$

As expected, $b_1 = b_2 = b$. But, $b_3 = b_4 = -.5\,h^2b$; i.e., the weight for the dam is $-.5\,h^2$ of that for the progeny. This weight is certainly different from weights of the historical daughter-dam comparison where:

Sire value = Daughter average - Dams' average.

With such a procedure, $b_2 = -b_1$ rather than $-.5\,h^2b_1$.

The equal parent or American index also weighted the dam record too much. The "logic" for the equal parent index was that

Progeny value = .5 (Sire value) + .5 (Dam value).

Rearrangement of the terms gives

Sire value = 2 (Progeny average value) - Dams' average,

so that $b_2 = -.5\, b_1$ rather than $b_2 = -.5\, h^2 b_1$.

The correct procedure can be simplified if the dams (sire's mates) are assumed to be unrelated so that only two b's are needed because each daughter record receives the same weight as any other daughter record and each dam record receives the same weight as any other dam record.

If X_1 is the average of single records of p daughters and

X_2 is the average of single records of the p dams,

the equations to find the weights are:

$$\left[\frac{1 + (p-1)\, a_{11} h^2}{p}\right] b_1 + \left[\frac{a_{12} h^2}{p}\right] b_2 = a_{1\alpha} h^2$$

$$\left[\frac{a_{12}\, h^2}{p}\right] b_1 + \left[\frac{1}{p}\right] b_2 = a_{2\alpha} h^2 .$$

The off-diagonal coefficient corresponds to the average covariance between the daughters and dams. Each daughter has covariance $a_{12} h^2 \sigma_X^2$ with her dam but a covariance of zero with the other p - 1 dams resulting in $\mathrm{Cov}(X_1, X_2) = a_{12} h^2 \sigma_x^2 / p$.

Usually $a_{11'} = .25$, $a_{12} = .5$, $a_{1\alpha} = .5$ and $a_{2\alpha} = 0$ so that

$$b_1 = \frac{2p}{p+\lambda-h^2} \text{ and } b_2 = -.5\, h^2 b_1.$$

Note the similarity to the b_1 when dams are not considered, e.g., if $h^2 = .25$,

$$b_1 = \frac{2p}{p+14.75} \text{ rather than } b_1 = \frac{2p}{p+15} \text{ when the dams' records are ignored.}$$

Similarly, the r_{TI} changes only slightly because $a_{2\alpha} = 0$. If $h^2 = .25$,

$$r_{TI} = \sqrt{\frac{p}{p + 14.75}} \text{ rather than } r_{TI} = \sqrt{\frac{p}{p + 15}} \text{ when dams are ignored.}$$

In other words, correction of a progeny proof for differences in mates does not increase accuracy of evaluation much. The advantage of correcting for differences in mates is to eliminate bias that would inflate proofs of sires mated to better than average dams.

PROGENY WITH DIFFERENT NUMBERS OF RECORDS

Often in evaluation of sires, progeny may have different numbers of records. A common example is that Standardbred trotting horses may have many more than one racing record. One solution to the problem of weighting these records is to set up one equation for each record. Then the correct weight would be found for each record, but many equations would be needed. If simplified equations are used, diagonal coefficients will be 1. RHS's will be $a_{i\alpha}h^2$ as before for all i. In the case of half sibs for sire evaluation, RHS's will all be $.5\ h^2$. Off-diagonal coefficients will be of two kinds. Coefficients corresponding to covariances among records on the same animal will be repeatability, r, because the covariance, $\sigma_{X'X} = r\sigma_X^2$. The other coefficients will be $a_{ij}h^2$ as before where

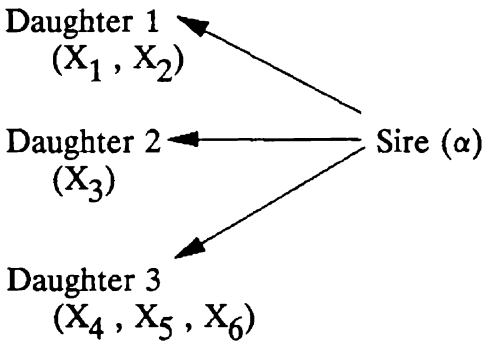
a_{ij} is the relationship between pairs of animals that made the records. In sire evaluation from paternal half-sib records, these coefficients will be .25 h^2.

Example: Daughter 1 has two records X_1 and X_2 ,

Daughter 2 has one record X_3 ,

Daughter 3 has three records X_4 , X_5 , and X_6 .

To estimate the additive genetic value of their sire from



the index will be: $I = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6$.

The equations to find the b's are:

$$b_1 + rb_2 + .25\,h^2b_3 + .25\,h^2b_4 + .25\,h^2b_5 + .25\,h^2b_6 = .5\,h^2$$

$$rb_1 + b_2 + .25\,h^2b_3 + .25\,h^2b_4 + .25\,h^2b_5 + .25\,h^2b_6 = .5\,h^2$$

$$.25\,h^2b_1 + .25\,h^2b_2 + b_3 + .25\,h^2b_4 + .25\,h^2b_5 + .25\,h^2b_6 = .5\,h^2$$

$$.25\,h^2b_1 + .25\,h^2b_2 + .25\,h^2b_3 + b_4 + rb_5 + rb_6 = .5\,h^2$$

$$.25\,h^2b_1 + .25\,h^2b_2 + .25\,h^2b_3 + rb_4 + b_5 + rb_6 = .5\,h^2$$

$$.25\,h^2b_1 + .25\,h^2b_2 + .25\,h^2b_3 + rb_4 + rb_5 + b_6 = .5\,h^2$$

An easier way to obtain the same result is to divide the daughters into groups with the same number of records for each daughter in a group:

X_1 is the average of p_1 daughters with $n_1 = 1$ record,

X_2 is the average of p_2 daughters with $n_2 = 2$ records,

$\vdots$

X_N is the average of p_N daughters with $n_N = N$ records.

The equations to find the weights for

$I = b_1X_1 + b_2X_2 + \cdots + b_NX_N$ are:

$$d_1b_1 + a_{12}h^2b_2 + \cdots + a_{1N}h^2b_N = a_{1\alpha}h^2$$

$$a_{12}h^2b_1 + d_2b_2 + \cdots + a_{2N}h^2b_N = a_{2\alpha}h^2$$

$$\vdots \qquad\qquad\qquad\qquad\qquad\qquad \vdots$$

$$a_{1N}h^2b_1 + a_{2N}h^2b_2 + \cdots + d_Nb_N = a_{N\alpha}h^2.$$

If all animals are half-sibs,

$$d_i = \frac{\frac{1 + (n_i-1)\,r}{n_i} + (p_i-1).25\,h^2}{p_i} \quad \text{with}$$

$$a_{ij}h^2 = .25\,h^2, \quad \text{and} \quad a_{i\alpha}h^2 = .5\,h^2.$$

The r_{TI} can be computed as usual as the square root of the sum of products of the b's and the corresponding additive relationships on the RHS.

EVALUATION WITH FULL SIB GROUPS

Some species such as swine and poultry may have full-sib progeny groups. Each male may be mated to more than one female. If each female produces only one set of progeny, the animals in each group will be related as full sibs ($a_{ii'} = .5$) but also will be related as paternal half sibs ($a_{ij} = .25$) to animals in other groups.

If p_i is the number in each full sib group, $n_i = 1$ and the sire is to be evaluated, ($a_{i\alpha} = .5$), the equations defining the b's are:

$$
\begin{aligned}
d_1 b_1 &+ .25\, h^2 b_2 + \cdots + .25\, h^2 b_N = .5\, h^2 \\
.25\, h^2 b_1 &+ d_2 b_2 + \cdots + .25\, h^2 b_N = .5\, h^2 \\
&\vdots \qquad\qquad\qquad\qquad\qquad\quad \vdots \\
.25\, h^2 b_1 &+ .25\, h^2 b_2 + \cdots + d_N b_N = .5\, h^2
\end{aligned}
\qquad \text{where}
$$

$$
d_i = \frac{1 + (p_i - 1)\, .5\, h^2}{p_i} .
$$

Modifications would, of course, have to be made for some $n_i > 1$, for other possible relationships such as maternal sibs and for any environmental correlation which is very likely for animals in the same litter as well as maternal effects in common as discussed in the chapter on imbedded traits.

Use of other combinations of relatives in the selection index is illustrated in problem sets. Often the animal will have records (one or more), progeny with records and relatives with records through both the paternal and maternal sides of the pedigree.

CHAPTER TEN

PROBABILITY STATEMENTS ABOUT TRUE VALUES

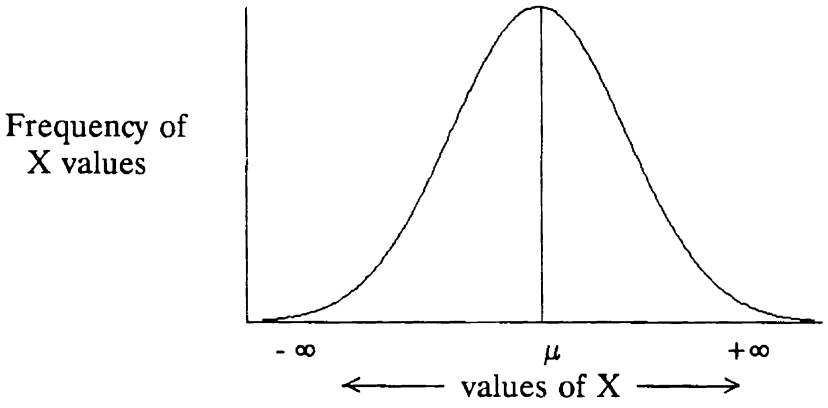
One property of selection index is that the average true value, T, for animals with the same index value, $I=I_0$, is I_0. Thus, I_0 is the mean of a subdistribution of T for animals with the same index, I_0, i.e., the distribution is conditional on I_0 and the accuracy of prediction of I, r_{TI}. The variance of T for $I=I_0$ depends on r_{TI} and σ_T^2 but not on I_0 :

$$\sigma^2_{T|I=I_0} = (1-r^2_{TI})\ \sigma^2_T .$$

If T and I follow a bivariate normal distribution, I_0 and $\sigma^2_{T|I=I_0}$ determine the distribution of T for $I=I_0$. After a review of the normal distribution, how to use the conditional mean and variance to make probability statements about T for $I=I_0$ will be described.

THE NORMAL DISTRIBUTION

The mean, μ, and the variance, σ^2, completely determine the normal distribution. The normal distribution follows the so-called bell shaped curve.



Let X be a set of values having a normal distribution. The mean is also the median of the X values, i.e., one-half the values of X are greater than μ and one-half the values are less than μ. The distribution of values is also symmetrical. The curve on the right-hand side of μ is the mirror image of the curve on the left-hand side of μ. The variance, σ^2, determines how flat or how peaked the curve is. A large σ^2 tends to flatten the curve and a small σ^2 tends to peak the values about μ. The total frequency of X's is 1 or 100%. Thus, the area under the normal curve is also 1. The fraction of the area above μ is 0.5 and the fraction below μ is also 0.5.

A table of areas under the normal curve describes the fraction of the area between μ and $\mu + t\sigma$ or equivalently between μ and $\mu - t\sigma$ because the distribution is symmetrical. This fraction corresponds to the probability that a random value of X will be between μ and $\mu + t\sigma$. The values of **t** are multipliers of the standard deviation.

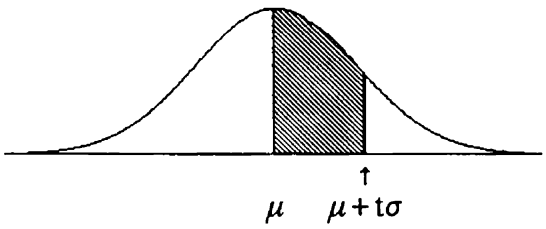
These are two uses of such a table (e.g., Table 10.1):

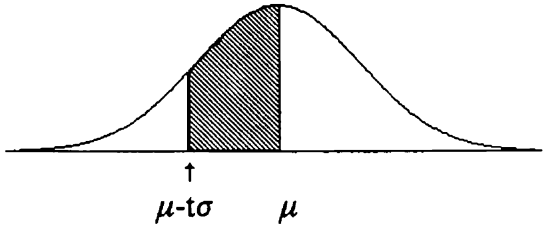
1) To find probabilities (fractions of total area) corresponding to truncation points which can be expressed as $\mu + t\sigma$ or $\mu - t\sigma$ depending on which side of μ the truncation point is located and
2) To find truncation points expressed as $\mu + t\sigma$ or $\mu - t\sigma$ corresponding to required probabilities.

Examples of Finding Probabilities Corresponding to Specified Truncation Points

Let $\sigma = 2$ and $\mu = 10$ for a distribution of values of X. The problem is to find the probability that a random value of X will be between 6 and 12.

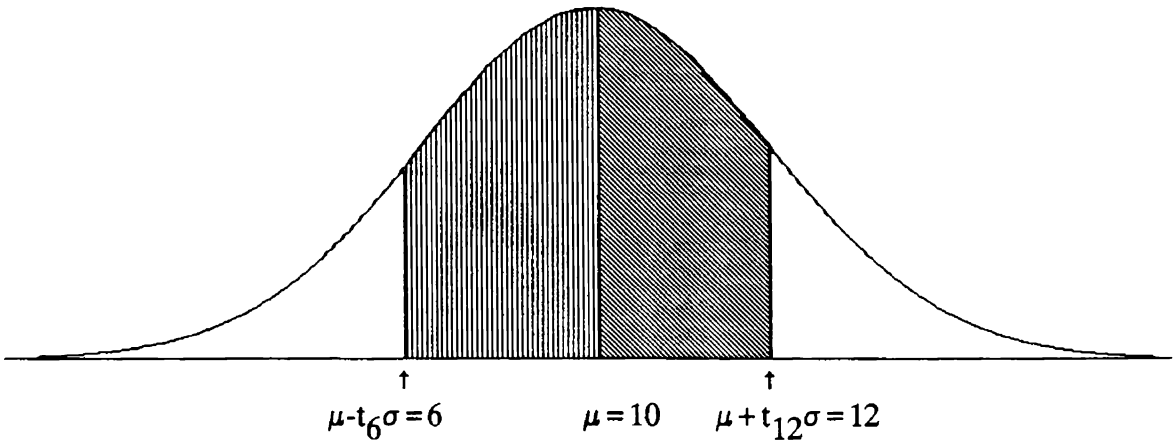
TABLE 10.1. AREAS UNDER THE NORMAL CURVE





Truncation Point t	Area between $\mu+t\sigma$ and μ or between $\mu-t\sigma$ and μ
.0	.0
.1	.04
.2	.08
.3	.12
.4	.16
.5	.19
.6	.23
.7	.26
.8	.29
.9	.32
1.0	.34
1.1	.36
1.2	.38
1.3	.40
1.4	.42
1.5	.43
1.6	.445
1.7	.455
1.8	.464
1.9	.471
2.0	.477
•	•
•	•
•	•
2.5	.494
•	•
•	•
•	•
3.0	.499

The first step in finding probabilities is to draw a picture to describe the problem as shown for this example. The picture usually will give an approximate answer as a way of checking the logic of the exact answer.



Because the table gives the area between μ and $\mu + t\sigma$, the solution is to find the area between $\mu = 10$ and $\mu + t_{12}\sigma = 12$ and the area between $\mu = 10$ and $\mu - t_6\sigma = 6$. Subscripts on the **t**'s are to identify the truncation points. Total area is sum of the two parts. In more formal terms: $P(6 < X < 12) = P(6 < X < 10) + P(10 < X < 12)$.

To use the table, t_{12} and t_6 must be calculated:

$\mu + t_{12}\sigma = 12$ but $\mu = 10$ and $\sigma = 2$. Thus $t_{12} = (12 - 10) / 2 = 1$.

The corresponding area (Table 10.1) is .34 .

The general method of finding a **t** corresponding to a positive truncation point, i.e., a point greater than the mean is to write the equality and solve for **t**:

$$t = \frac{\mu + t\sigma - \mu}{\sigma} = \frac{\text{truncation point} - \mu}{\sigma} .$$

On the left side of μ, $\mu - t_6\sigma = 6$ and thus, $10 - t_6 (2) = 6$ and $t_6 = 2$.

For $t = 2$, the corresponding area is .477 between 6 and 10.

For t corresponding to truncation points less than the mean,

$t = -\dfrac{\text{truncation point} - \mu}{\sigma}$. In general, $t = \dfrac{|\text{truncation point} - \mu|}{\sigma}$.

The total area between 6 and 12 is .477 + .34 = .817, which is the fraction of X's expected to have values between 6 and 12 or equivalently the probability that any random X will be between 6 and 12.

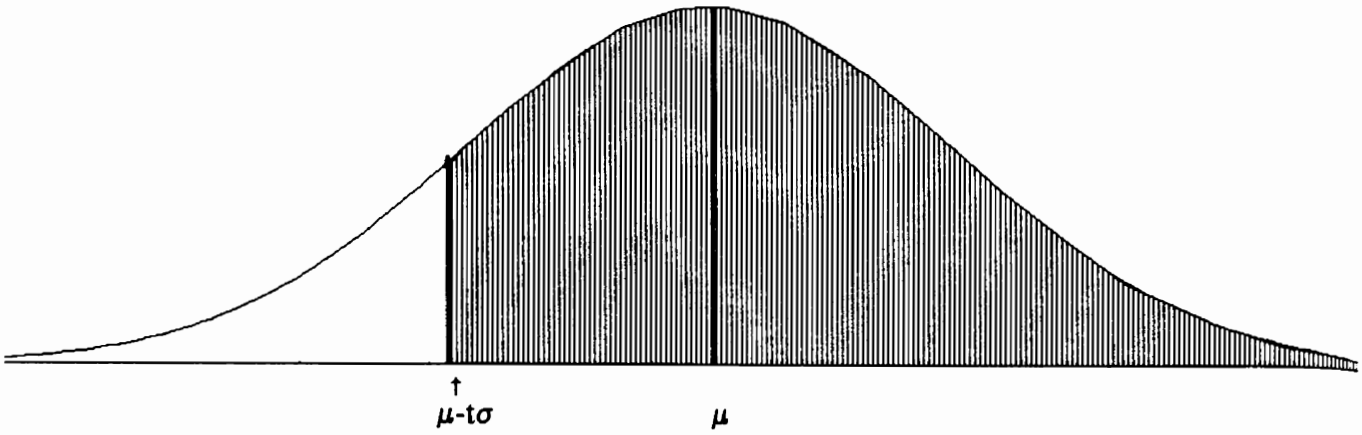
Another example is to find the probability of an X value above a truncation point, such as 12. Then, $\mu + t\sigma = 12$ with $\mu = 10$ and $\sigma = 2$ and $P(10 < X < 12) = .34$. Obviously, $P(10 < X < \infty) = .5$. Thus, $P(12 < X < \infty) = .5 - .34 = .16$.

The probability of a random X less than 12 can be found by similar logic, i.e., $P(-\infty < X < 12) = P(-\infty < X < 10) + P(10 < X < 12) = .50 + .34 = .84$.

Examples of Finding Truncation Points Corresponding to Specified Probabilities

Find the region which includes 90% of values of X which is also the probability that a random value of X will be in that region. These ranges may be chosen so that they are symmetrical about μ, i.e., $\mu + t\sigma$ is the upper limit, and $\mu - t\sigma$ is the lower limit with t the same in both upper and lower limits. First, draw the picture which will show that the area from μ to $\mu + t\sigma$ must be .90/2 = .45. The t corresponding to an area of .45 is about halfway between 1.6 and 1.7 so let t = 1.65. If $\mu = 10$ and $\sigma = 2$ as before, the upper limit is 10 + 1.65(2) = 13.30, and the lower limit is 10 - 1.65(2) = 6.70 .

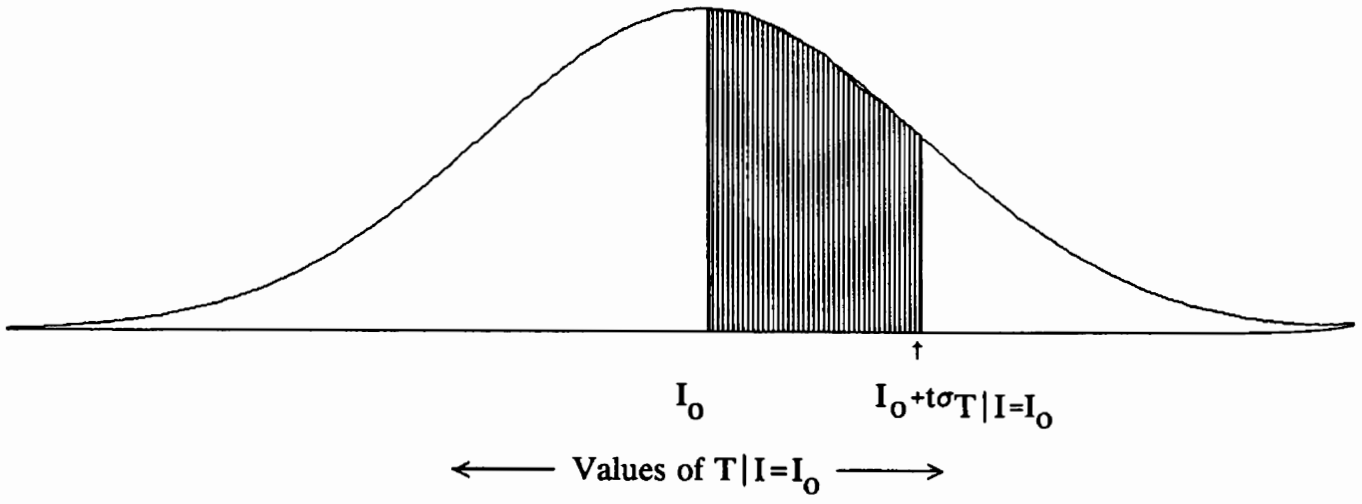
Next, find the truncation point which 90% of the values of X will exceed for the example with $\mu = 10$ and $\sigma = 2$.



The t corresponding to an area of .40 between μ and μ - tσ must be found. From Table 10.1, t = 1.3. Thus, the truncation point is 10 - 1.3(2) = 7.4. The probability of a random X having a higher value is 90%. Also, 90% of the values of X will be greater than 7.4.

APPLICATIONS TO ESTIMATING TRUE VALUE

One property of the selection index is $\mu_{T|I=I_0} = I_o$. Thus, I_0 corresponds to the mean of the distribution of T values for animals with the same index. Thus, I_0 can be substituted for μ of the normal distribution. Similarly, $\sigma_{T|I=I_0}$ will be substituted for σ of the general distribution. A typical picture is:



EXAMPLES

Probability Statements About Additive Genetic Values; $T = G_{A_\alpha}$.

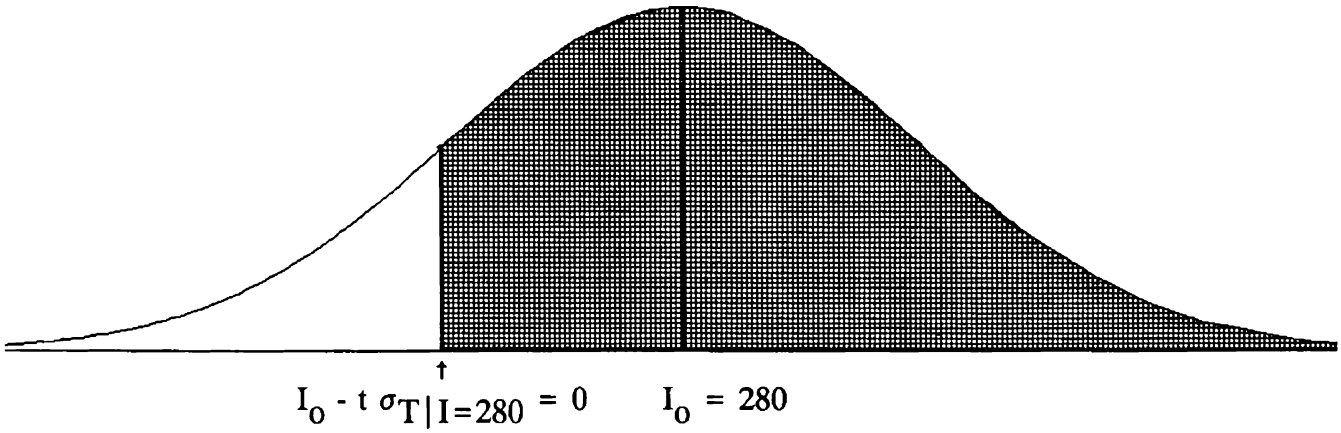
A bull has 35 progeny with 1 record each averaging +200; $h^2 = .25$; $\sigma^2_{10} = 1000000$. What is the probability his true additive genetic value is greater than zero?

For $h^2 = .25$, $\lambda = 15$ so that

$$b = \frac{2p}{p+15} = \frac{70}{50}, \quad I_0 = (1.4)(200) = 280, \text{ and } r^2_{TI} = \frac{p}{p+15} = .70.$$

Thus $\sigma^2_{T|I=I_0} = (1 - .70)(1000000) = 300000$ and $\sigma_{T|I=I_0} = 548$.

The picture is:



Then, $t = \frac{|0 - 280|}{548} = .51$. The corresponding area gives the fraction between 0 and 280 as .19. Thus, the probability that T for the bull will exceed zero is .50 + .19 = .69. Correspondingly, the probability that his true value is less than zero is 1.00 - .69 = .31.

Probability Statements About A Future Record

The previous discussion was about the probability that an animal's additive genetic value was between, above, or below certain truncation points given the index estimate of

breeding value and the corresponding r_{TI} and σ^2_G. In fact, if the animal actually makes a record, in addition to its genetic value, a new random environmental effect influences the record. Thus, variance of records for animals with a predicted genetic value depends on the variance of additive genetic values given the index plus the variance of environmental effects. In this case of predicting a future record, $T = X_\alpha = G_{A_\alpha} + E_\alpha$, I is the prediction of a future phenotypic record of animal α that has no previous record. In this example, the assumption is that $G_\alpha = G_{A_\alpha}$. The selection index equations to find the appropriate weights for the X's are, as usual, on the left-hand sides, the variances and covariances of the X's. The right hand sides are:

$$\sigma_{X_iT} = E\left[(X_i)(G_{A_\alpha} + E_\alpha)\right] = E\left[(G_{A_i} + E_i)(G_{A_\alpha} + E_\alpha)\right]$$
$$= E(G_{A_i}G_{A_\alpha} + G_{A_i}E_\alpha + G_{A_\alpha}E_i + E_iE_\alpha).$$

The middle two terms are genetic-environmental covariances which are usually assumed to be zero. The first term is $a_{i\alpha}\sigma^2_{10}$ for $G_A = G$, and the last term is the covariance between environmental effects on a record of i and on a record of α which may or not be zero. With no environmental covariance, the right-hand sides are $a_{i\alpha}\sigma^2_{10} = a_{i\alpha}h^2\sigma^2_X$ as for predicting additive genetic value so that the index for predicting a future record is exactly the same as for predicting additive genetic value. The reason is that there is no way of predicting a random and independent E_α for the new record.

The r_{TI} and σ^2_T, however, are different from the case when I predicted G_{A_α}. Now $\sigma^2_T = E(T^2) = E(G_{A_\alpha} + E_\alpha)^2 = \sigma^2_{G_A} + \sigma^2_E = \sigma^2_X$ rather than $\sigma^2_{G_A} = h^2\sigma^2_X$. The numerator of r_{TI} is $\Sigma b_i a_{i\alpha} h^2 \sigma^2_X$ as before, but

$$r_{TI} = \sqrt{\frac{\Sigma b_i a_{i\alpha} h^2 \sigma^2_X}{\sigma^2_X}} = \sqrt{h^2 \Sigma b_i a_{i\alpha}}$$

rather than $\sqrt{\Sigma b_i a_{i\alpha}}$ because prediction of E_α is zero. Then $\sigma^2_{T|I=I_0} = (1-h^2\Sigma b_i a_{i\alpha})\sigma^2_X$ rather than $(1-\Sigma b_i a_{i\alpha})h^2\sigma^2_X$. Notice that many of the same quantities, $\Sigma b_i a_{i\alpha}$, h^2, σ^2_X, are involved whether prediction is for G_{A_α} or $X_\alpha = G_{A_\alpha} + E_\alpha$; the arrangement, however, is different in important ways.

Prediction of a Progeny Record from Prediction of Additive Genetic Values of the Parents

The application of these distributional properties makes sense primarily when records of ancestors are used in estimating the animal's genetic value, as for example, if the sire's and dam's estimated additive genetic values are used in estimating the additive genetic value of their progeny: $\hat{G}_{progeny} = \frac{\hat{G}_{sire} + \hat{G}_{dam}}{2}$ which also predicts a record of the progeny.

The r^2_{TI} for additive genetic value of a progeny equals one-fourth the sum of the r^2_{TI} for additive genetic value of sire and dam. This equality can be shown by setting up the equations to predict the additive genetic values of sire and dam and then to predict the average of the additive genetic values of the sire and dam. Assume for milk yield that $h^2 = .25$ and $\sigma^2_X = (2000 \text{ lb})^2$.

The following table shows the effect of increasing r^2_{TI} for sire and dam on $\sigma_{X|I=I_0}$ of the progeny.

Sire r^2_{TI}	Dam r^2_{TI}	Progeny	$\sigma_{X\|I=I_0}$ (for progeny)	95% probability range = $I_0 \pm 1.96\ \sigma_{X\|I=I_0}$
0	0	0	2000 lb.	$I_0 \pm 3920$
.25	0	.0625	1984	$I_0 \pm 3889$
.25	.25	.1250	1968	$I_0 \pm 3857$
.50	.50	.2500	1936	$I_0 \pm 3795$
.75	.50	.3125	1920	$I_0 \pm 3763$
.75	.75	.3750	1904	$I_0 \pm 3732$
1.00	.75	.4375	1887	$I_0 \pm 3699$
1.00	1.00	.5000	1871	$I_0 \pm 3667$

The obvious conclusion from this chart is that the average error of predicting a record of a progeny ($\sigma_{X|I=I_0}$) does not decrease very much even with perfect prediction of the parents' genetic values when σ^2_E is relatively large.

Probability Statements About Differences in Genetic Values for Unrelated Animals

Assume animal 1 has index value I_1 with $r^2_{TI_1}$ and animal 2 has index value I_2 with $r^2_{TI_2}$. Differences in true additive genetic values for animals with index values I_1 and I_2 will have a distribution corresponding to the definition of $T = G_{A_1} - G_{A_2}$. The immediate problem is to determine the mean and variance for the distribution of $T_1\text{-}T_2|I_1\text{-}I_2$. The mean is the same as the mean of $(T_1|I = I_1)\text{-}(T_2|I = I_2)$ and is

$$E[(T_1|I=I_1) - (T_2|I=I_2)] = E(T_1|I=I_1) - E(T_2|I=I_2) = I_1 - I_2 .$$

The variance can be determined by the rules for the variance of a linear function or can be derived from the selection index procedure for

$T = G_{A_1} - G_{A_2}$ with $I = b_1X_1 + b_2X_2$. By rules for variance of a linear function,

$$V[(T_1|I=I_1) - (T_2|I=I_2)] = V(T_1|I=I_1) + V(T_2|I=I_2)$$

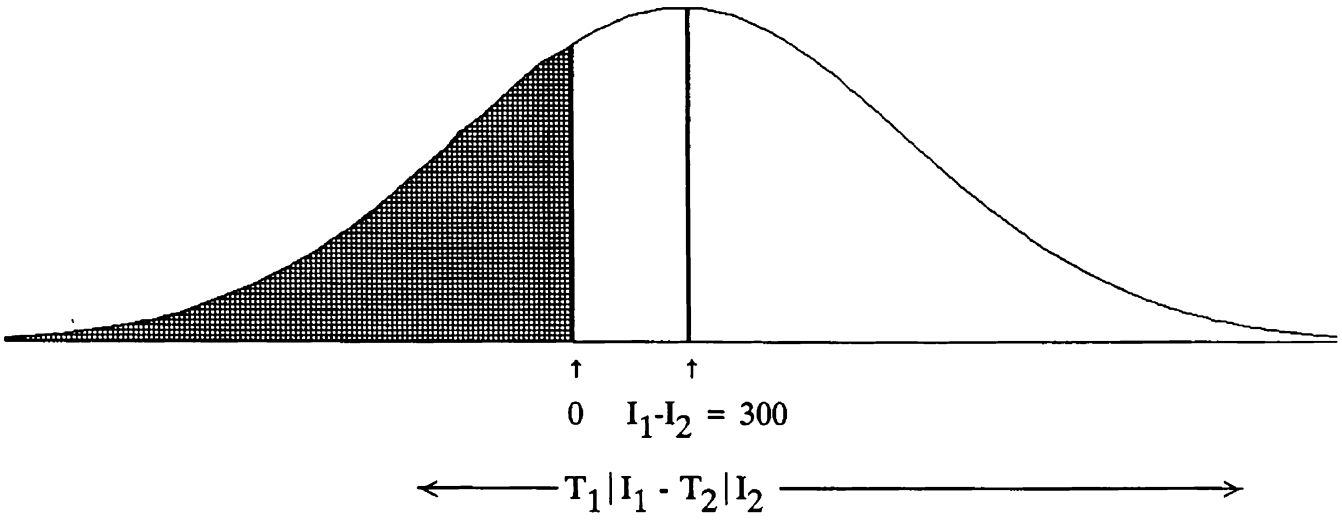
because the covariance between $T_1|I=I_1$ and $T_2|I=I_2$ is zero if the records in I_1 and I_2 are independent. Thus,

$$V[(T_1|I=I_1) - (T_2|I=I_2)] = (2 - r^2_{TI_1} - r^2_{TI_2})\sigma^2_G.$$

These parameters can be used to make probability statements about the difference in true values for animals with indexes I_1 and I_2. In this case, I_1-I_2 replaces μ of the general discussion and $(2 - r^2_{TI_1} - r^2_{TI_2})\sigma^2_G$ replaces σ^2.

As a numerical example, suppose $I_1 = 500$ and $I_2 = 200$, i.e., I_1-$I_2 = 300$ and $r^2_{TI_1} = .75$ and $r^2_{TI_2} = .25$ and $\sigma^2_G = (1000)^2$. What is the probability that the true difference in genetic values is zero or less, i.e., that an animal with $I_2 = 200$ actually has equal or greater true value than an animal with $I_1 = 500$?

The picture is:



The problem is to find $\text{Prob}[(T_1 \mid I_1 - T_2 \mid I_2) \leq 0]$.

Thus,

$$0 = (I_1 - I_2) - t\sqrt{(2 - r^2_{TI_1} - r^2_{TI_2})\sigma^2_G} \; .$$

Then,

$$t = (300 - 0)/[\sqrt{(2 - .75 - .25)}\,(1000)] = .3 .$$

The corresponding area between 0 and 300 is .12 and the area below 0 is .5 - .12 = .38, which is the probability that the animal with the smaller index, $I_2 = 200$, will have a higher additive genetic value than the animal with the larger index, $I_1 = 500$.

A more direct approach would be to define $T = G_{A_1} - G_{A_2}$, use all information to predict T, and then follow the general selection index procedure.

SUMMARY OF DISTRIBUTIONS ASSOCIATED WITH SELECTION INDEX

Often there is some confusion about what μ and σ are. Actually neither μ nor σ has any meaning unless defined in terms of the variable they describe. In development of the selection index for a particular trait there are at least 6 variables.

1) The basic distribution is of phenotypic records, the P's, or as they have also been called, the X's. The mean is μ_X and the variance is $\sigma^2_X = \sigma^2_G + \sigma^2_E$ (the genetic plus environmental variance).

2) For additive genetic values, the mean is $\mu_G = 0$ and the variance is $\sigma^2_{10} = h^2\sigma^2_X$.

3) The criterion for predicting a G is the index estimate, I. The mean is $\mu_I = 0$ and the variance is $\sigma^2_I = r^2_{TI}\sigma^2_G$. Note that $\sigma^2_I \leq \sigma^2_G$ because $r^2_{TI} \leq 1$.

4) Animals with the same index value may not have the same true value. The distribution of true values given an index value has mean $\mu_{T|I=I_o} = I_o$ and variance $\sigma^2_{T|I=I_o} = (1-r^2_{TI})\sigma^2_T ; [(1-r^2_{TI})\sigma^2_G$, if T=G].

5) Records of an animal with an index value I_o have a different distribution from records with no estimate of true value. The distribution of records for animals with an index of I_o has mean $\mu_{X|I=I_o} = I_o$ and variance $\sigma^2_{X|I=I_o} = (1-r^2_{TI})\sigma^2_X$ when r^2_{TI} is for predicting a future record, X.

6) The difference in additive genetic values for animals with index values I_1 and I_2 is distributed with mean $\mu_{T_1|I=I_1} - \mu_{T_2|I=I_2}$ and variance $\sigma^2_{T|I=I_1} + \sigma^2_{T|I=I_2} = (2-r^2_{TI_1}-r^2_{TI_2})\sigma^2_G$ when the indexes are independent.

The standard deviations for the six distributions are:

1) X , $\sigma_X = \sqrt{\sigma_G^2 + \sigma_E^2}$;

2) G_A , $\sigma_G = \sqrt{h^2}\sqrt{\sigma_G^2 + \sigma_E^2}$;

3) I to predict G_A , $\sigma_I = r_{TI}\sigma_G$;

4) G given $I=I_o$, $\sigma_{T|I=I_o} = \sqrt{(1-r_{TI}^2)\ \sigma_G^2}$;

5) X given $I=I_o$ when r_{TI}^2 is for predicting X, $\sigma_{X|I=I_o} = \sqrt{(1-r_{TI}^2)\ \sigma_X^2}$

6) $G_1|I=I_1 - G_2|I=I_2$, $\sigma_{G_1-G_2|I_1-I_2} = \sqrt{(2-r_{TI_1}^2-r_{TI_2}^2)\ \sigma_G^2}$.

CHAPTER 11

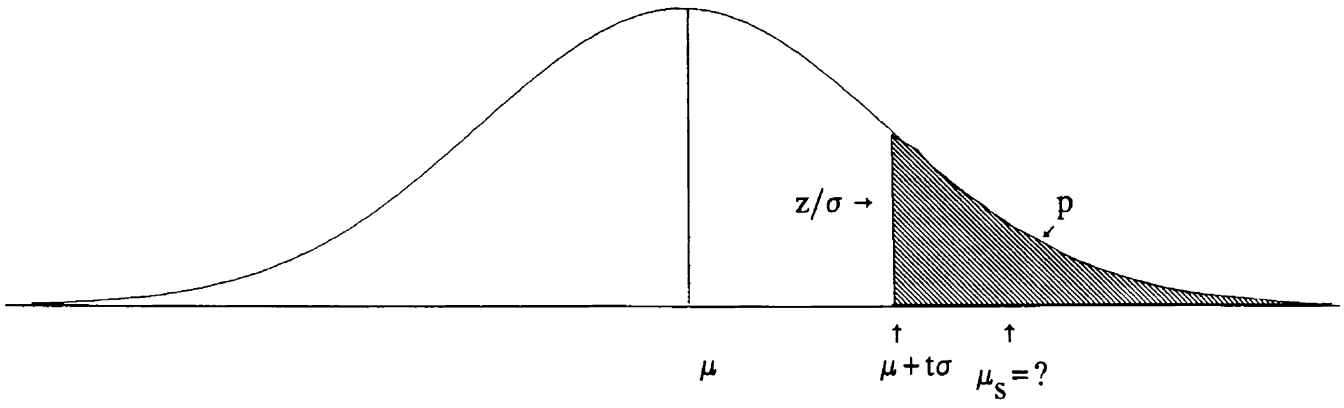
SUPERIORITY OF SELECTED GROUPS

AVERAGE OF SELECTED GROUP

The basic principle in selection is to select the best and cull the rest. The selection index is the best linear method of evaluating animals to determine which to select or cull. The selection index is unbiased so an estimate of the superiority of the selected group is simply the average index of the selected group minus the average index of the whole group from which the selected group came. Another question, however, is how to determine how much better the selected ones are expected to be than the original group before the indexes are calculated? The answer to this question relies on theory based on the normal distribution of true values and index values.

THE NORMAL DISTRIBUTION

The basic problem is this. If a fraction, **p**, is selected from a normal distribution with mean, μ, and variance, σ^2, what will be the mean of the selected group, μ_s ? The problem may be diagrammed as:



The truncation point, $\mu + t\sigma$, depends on **p** as in Chapter 10.

The expected or average value of the fraction **p** can be found from integrating:

$$\mu_S = \frac{1}{p} \int_{\mu+t\sigma}^{\infty} xf(x)dx = \mu + D\sigma \text{ where } f(x) \text{ is the density function}$$

of the normal distribution. Fortunately, tables are available that give values of D given the fraction selected, **p**.

The difference from μ, $D\sigma$, is $z\sigma/p$ where z/σ is the height of the normal curve at the truncation point and **p** is the fraction selected. Note that $D = z/p$, the height of the normal curve for $\sigma = 1$. The tables of D are based on the normal distribution with $\sigma = 1$. To convert the table of D values to any other distribution, multiply by σ. Many texts use **i** rather than D to describe the standardized selection differential.

Note that $\mu_S - \mu = D\sigma$, which is sometimes known as the selection differential (not standardized for σ). If $\mu = 0$, $\mu_S = D\sigma$.

The table of D for small samples is based on expected values of order statistics (Table 11.1). The values are not the same as z/**p**. The table of D for large samples is the same as z/**p** (Table 11.2). Dr. C. R. Henderson proposed an approximate correction for sample size for this table, i.e., $D' = D - \frac{.25}{s}$, where **s** is the number selected. Note **s** is *not* the number available for selection.

EXAMPLE OF SELECTION BIAS

A breed organization reports a dairy bull has 100 daughters. The average of the top 20 is +1000 lb of milk. The standard deviation of records of cows by the same sire is about 2000 lb. What would the average of the 100 daughters be expected to be?

The fraction selected is 20 of 100 or, 20%. The corresponding value of D = 1.4. Thus, D' = 1.4 - .25/20 = 1.3875.

From the normal theory

$\mu_S = \mu + D\sigma$, i.e.,

1000 = μ + 1.3875(2000) so that μ = 1000 - 1.3875(2000) = -1775 lb .

Evaluation of the bull on his top 20 daughters would have been considerably misleading. An interesting question is what should be the number of daughters to use in the formula for estimating the genetic value of this bull--20 or 100 or something else? Schaeffer et al. (1970) developed a solution for this problem which depends mainly on the fraction selected.

TABLE 11.1. EXPECTED AVERAGE OF A GROUP SELECTED OUT OF A SAMPLE FROM A NORMAL POPULATION WHEN THE SAMPLE SIZE IS SMALL (IN UNITS OF σ = 1)

Sample Size	Number selected													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	.56	.00												
3	.85	.42	.00											
4	1.03	.66	.34	.00										
5	1.16	.83	.55	.29	.00									
6	1.27	.95	.70	.48	.25	.00								
7	1.35	1.05	.82	.62	.42	.23	.00							
8	1.42	1.14	.92	.73	.55	.38	.20	.00						
9	1.49	1.21	1.00	.82	.65	.50	.35	.19	.00					
10	1.54	1.27	1.07	.89	.74	.60	.46	.32	.17	.00				
11	1.59	1.32	1.12	.96	.81	.68	.55	.42	.29	.16	.00			
12	1.63	1.37	1.18	1.02	.88	.75	.63	.51	.39	.27	.14	.00		
13	1.67	1.42	1.23	1.07	.93	.81	.69	.58	.48	.37	.26	.14	.00	
14	1.70	1.46	1.27	1.12	.99	.87	.76	.65	.55	.45	.35	.24	.13	.00
15	1.74	1.49	1.31	1.16	1.03	.92	.81	.71	.61	.52	.42	.33	.23	.12

TABLE 11.2. EXPECTED AVERAGE OF CERTAIN FRACTIONS SELECTED OUT OF A SAMPLE FROM A NORMAL POPULATION (IN UNITS OF σ = 1)

Table for .001-.099 Selected

	.000	.001	.002	.003	.004	.005	.006	.007	.008	.009
.00		3.400	3.200	3.033	2.975	2.900	2.850	2.800	2.738	2.706
.01	2.660	2.636	2.600	2.569	2.550	2.527	2.500	2.482	2.456	2.442
.02	2.420	2.400	2.386	2.370	2.363	2.336	2.323	2.311	2.293	2.283
.03	2.270	2.258	2.241	2.230	2.221	2.209	2.200	2.186	2.174	2.164
.04	2.153	2.146	2.136	2.126	2.116	2.107	2.098	2.087	2.079	2.071
.05	2.064	2.057	2.048	2.040	2.031	2.022	2.016	2.009	2.000	1.990
.06	1.985	1.977	1.971	1.965	1.958	1.951	1.944	1.937	1.931	1.925
.07	1.919	1.911	1.906	1.900	1.893	1.888	1.882	1.875	1.871	1.863
.08	1.858	1.852	1.846	1.841	1.837	1.834	1.826	1.820	1.815	1.810
.09	1.806	1.799	1.793	1.788	1.784	1.780	1.775	1.770	1.765	1.760

Table for .10-.99 Selected

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.10	1.755	1.709	1.667	1.628	1.590	1.554	1.521	1.488	1.458	1.428
.20	1.400	1.372	1.346	1.320	1.295	1.271	1.248	1.225	1.202	1.180
.30	1.159	1.138	1.118	1.097	1.078	1.058	1.039	1.021	1.002	.984
.40	.966	.948	.931	.913	.896	.880	.863	.846	.830	.814
.50	.798	.782	.766	.751	.735	.720	.704	.689	.674	.659
.60	.644	.629	.614	.599	.585	.570	.555	.540	.526	.511
.70	.497	.482	.468	.453	.438	.424	.409	.394	.380	.365
.80	.350	.335	.320	.305	.290	.274	.259	.243	.227	.211
.90	.195	.179	.162	.144	.127	.109	.090	.070	.049	.027

If the number selected is less than 500, subtract from D the quantity .25/s, where s is the number selected.

GENETIC SUPERIORITY OF SELECTED GROUP

A fraction of animals is to be selected for T based on their index values. What is the expected superiority in T of the selected group?

The selected I's, if normally distributed, will be expected to average $\mu_{I_S} = \mu_I + D\sigma_I$. Note that $\mu_I = 0 = \mu_T$ before selection, $\sigma_I = r_{TI}\sigma_T$, and $\mu_{T_S} = \mu_{I_S}$ because I is unbiased. Then, making these substitutions, $\mu_{T_S} = \mu_T + r_{TI}\sigma_T D$ as described in most animal breeding literature. The same result can be obtained by the regression of T on I:

$$\mu_{T_S} = \mu_T + b_{T\cdot I}(\mu_{I_S} - \mu_I) = \mu_T + (\sigma_{TI}/\sigma_I^2)(\mu_I + D\sigma_I - \mu_I)$$

$$= \mu_T + (\sigma_{TI}/\sigma_I)D = \mu_T + r_{TI}\sigma_T D \quad \text{by multiplying by } \sigma_T/\sigma_T.$$

Thus, the genetic selection differential per generation will be $\Delta G = r_{TI}D\sigma_T$. If L is the generation interval in years, then genetic progress per year,

$$\Delta G/yr = r_{TI}D\sigma_T/L.$$

For any given set of animals, however, the best estimate of the genetic superiority of the selected group is $\mu_{I_S} - \mu_I$, the difference in average index value of the selected and whole population. The indexes are unbiased predictions of genetic value so that averages of these are also unbiased. In fact, the difference in the averages is the selection index prediction of the difference between the selected group and the group from which they were selected.

The expression $\Delta G/yr = r_{TI}D\sigma_T/L$ can be used to compare the potential of various selection programs. This equation is the key equation for designing breeding programs for genetic improvement. Sometimes the best balance of r_{TI}, D, and L will have to be found.

Example of Finding Optimum Number of Progeny Per Sire and Number of Sires to Sample

Suppose that only 1000 progeny are available each year for progeny testing. Two replacements are needed each year from the males that are progeny tested. Assume $h^2 = .25$ and $\sigma_T = 1000$ lb. milk. The following table illustrates that neither the largest r_{TI} nor the greatest selection intensity gives the highest genetic progress.

SOME POSSIBLE COMBINATIONS OF NUMBER OF MALES PER SAMPLE AND NUMBER OF PROGENY PER SIRE

Number selected / Number sampled	%	Number progeny per male sampled	$r_{TI} = \sqrt{\frac{p}{p+15}}$	$D' = D - \frac{.25}{2}$	σ_T	ΔG
2 of 2	100	500	.985	0	1000 lb	0 lb
2 of 5	40	200	.964	.84	1000	810
2 of 20	10	50	.877	1.63	1000	1429
2 of 50	4	20	.756	2.03	1000	1535
2 of 100	2	10	.633	2.30	1000	1456
2 of 200	1	5	.500	2.54	1000	1270

Of the six combinations, testing 50 males each with 20 progeny seems to be best. In actual practice, income and cost values must be assigned to each plan. Since ΔG for selecting 2 of 20 sampled is nearly as great as ΔG for 2 of 50, that may be the most profitable plan. Other factors should also be considered in finding an optimum plan. The fraction of the population devoted to progeny proving is another variable in some cases. The generation interval may also be important.

The preceding example ignored the fact that ΔG is usually different for males and females since r_{TI}, D, and generation interval all may be different for males and females. Total expected genetic response per year depends on both as will be seen, although the expected genetic superiority of the offspring is the average of the superiorities of the selected males and females.

GENETIC VALUE OF PROGENY

Let $\Delta S = r_{TI_S} D_S \sigma_G$, where ΔS is the genetic superiority of selected sires, r_{TI_S} is the accuracy of the index for sires, and D_S is the selection intensity factor for sire selection. Similarly, let $\Delta D = r_{TI_D} D_D \sigma_G$, the genetic superiority of selected dams. Then, because progeny receive a sample half of the genetic value of each of their parents, the superiority of progeny as compared to randomly mating males and females is:

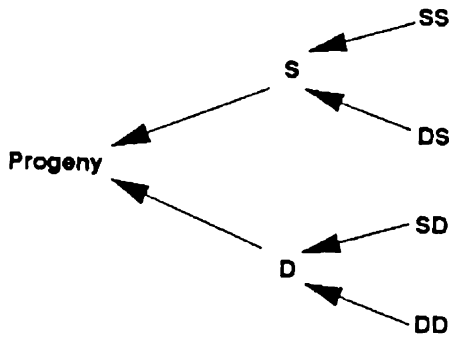
$$G_{progeny} = (\Delta S + \Delta D)/2.$$

GENETIC IMPROVEMENT PER YEAR

Let Δg be genetic improvement per year, L_S be the generation interval in years for sires, and L_D be the generation interval for dams. Then, $\Delta g = (\Delta S + \Delta D)/(L_S + L_D)$, which is not $[(\Delta S/L_S) + (\Delta D/L_D)]/2$ unless $L_S = L_D$. The proof is somewhat circular:

Let S be the genetic value of sires selected to produce the next generation and D be the value of selected dams. These selected sires are born L_S years before they produce replacement progeny with genetic value P. The genetic average of sires born L_S years ago is $P - L_S\Delta g$. The superiority of the selected sires over that average is ΔS. Thus, $S = P - L_S\Delta g + \Delta S$. Similarly, $D = P - L_D\Delta g + \Delta D$. Because $P = (S + D)/2$, then by substitution: $P = (S + D)/2 = (1/2)(P - L_S\Delta g + \Delta S + P - L_D\Delta g + AD)$. After subtracting P from both sides, $0 = -L_S\Delta g - L_D\Delta g + \Delta S + \Delta D$. Rearranging gives $\Delta g(L_S + L_D) = \Delta S + \Delta D$, and finally: $\Delta g = (\Delta S + \Delta D)/(L_S + L_D)$, a result due to Dickerson and Hazel (1944).

Rendel and Robertson (1950) extended this procedure to consider four paths of selection: sires of sires (SS), dams of sires (DS), sires of dams (SD), and dams of dams (DD) with generation intervals L_{SS}, L_{DS}, L_{SD}, and L_{DD}, respectively.



Let ΔSS, ΔDS, ΔSD, and ΔDD be the respective genetic superiorities of the selected grandparents as differences from their generation averages. For example, $\Delta SS = r_{TI_{SS}} D_{SS} \sigma_G$. By similar reasoning as before

$$SS = S - L_{SS}\Delta g + \Delta SS ,$$

$$DS = S - L_{DS}\Delta g + \Delta DS ,$$

$$SD = D - L_{SD}\Delta g + \Delta SD , \quad \text{and}$$

$$DD = D - L_{DD}\Delta g + \Delta DD .$$

Because $S = (SS + DS)/2$ and $D = (SD + DD)/2$, then

$G_{progeny} = (S + D)/2 = (SS + DS + SD + DD)/4$. Thus, by substitution,

$$\frac{S + D}{2} = (S - L_{SS}\Delta g + \Delta SS + S - L_{DS}\Delta g + \Delta DS + D - L_{SD}\Delta g + \Delta SD + D - L_{DD}\Delta g + \Delta DD)/4.$$

After rearranging and subtracting $(S + D)/2$ from both sides,

$\Delta g(L_{SS} + L_{DS} + L_{SD} + L_{DD}) = \Delta SS + \Delta DS + \Delta SD + \Delta DD$, so that

$\Delta g = (\Delta SS + \Delta DS + \Delta SD + \Delta DD)/(L_{SS} + L_{DS} + L_{SD} + L_{DD})$.

Genetic progress per year, then, is equivalent to the average superiority of the selected grandparents divided by the average generation interval of the different grandparent paths.

This expression or the preceding one involving just sires and dams can be used to compare expected genetic progress for different selection programs considering differences in generation intervals, selection intensities, and accuracies of prediction.

CHAPTER 12

SELECTION INDEX FLOW CHART FOR SINGLE TRAITS

The following six steps are a guide to using selection index for prediction of true value and design of optimum breeding programs. The various distributions involved with selection index properties and their means and variances also are described.

1) Define T.

2) $\hat{T} = I = b_1X_1 + \cdots + b_NX_N$, X's are available records, $I = b'x$.

3) Selection index equations determine b's which minimize $E[(T-I)^2]$ or maximize r_{TI}.

$$\begin{aligned} \sigma^2_{X_1} b_1 + \sigma_{X_1X_2} b_2 + \cdots &= \sigma_{X_1T} \\ \sigma_{X_1X_2} b_1 + \sigma^2_{X_2} b_2 + \cdots &= \sigma_{X_2T}, \\ \vdots \quad\quad \vdots \quad\quad &\quad\quad \vdots \end{aligned}$$

In matrix notation: $Pb = c$, so that $b = P^{-1}c$.

The $\sigma^2_{X_i}$, $\sigma_{X_iX_j}$, and σ_{X_iT} are determined from expected values, definition of T and models for X_i .

Models: $X_i = G_i + E_i$ or

$X_{ij} = G_i + PE_i + TE_{ij}$ for traits with repeated records.

$E(G_iG_j) = a_{ij}\sigma^2_{10} + a^2_{ij}\sigma^2_{20} + d_{ij}\sigma^2_{01} + a_{ij}d_{ij}\sigma^2_{11} + \cdots$

$E(E_iE_j) = c_{ij}\sigma^2_X$

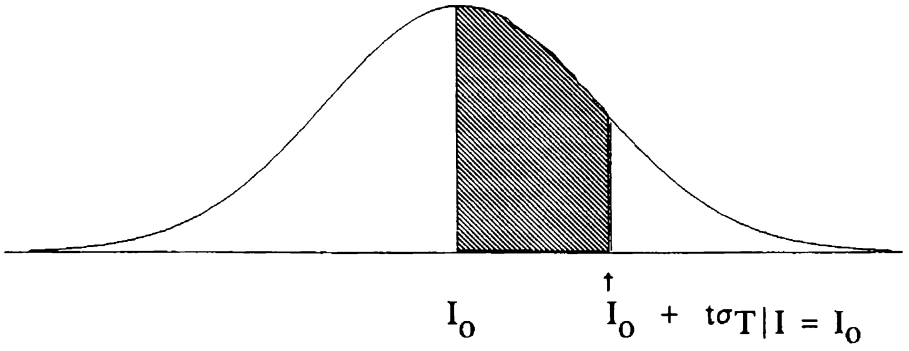
4) Rank animals using b_i's and X_i's, $I = b'x$ with actual X's.

5) Probability statements when $T|I=I_o$ has a normal distribution are based on:

$E(T|I = I_o) = I_o$

$V(T|I = I_o) = (1-r_{TI}^2)\sigma_T^2$

$r_{TI}^2 = (\Sigma b_i \sigma_{X_i T})/\sigma_T^2$; $b'c/\sigma_T^2$.



6) Theoretical comparison of selection programs

One path;

$$\Delta T = D\sigma_I = Dr_{TI}\sigma_T$$

Two paths, additive genetic value;

$$\Delta g/yr = \frac{\Delta S + \Delta D}{L_S + L_D} \quad \text{where } \Delta S = r_{TI_S} D_S \sigma_G, \text{ etc.}$$

Four paths, additive genetic value;

$$\Delta g/yr = \frac{\Delta SS + \Delta DS + \Delta SD + \Delta DD}{L_{SS} + L_{DS} + L_{SD} + L_{DD}}$$

DISTRIBUTIONS INVOLVED WITH PREDICTING ADDITIVE GENETIC VALUE

Distribution	Mean	Variance	
1) P_i, phenotypic	μ_p	σ^2_X	
2) X_i, adjusted phenotypic record, $P_i - \mu_p$	0	σ^2_X	
3) G_{A_i}, additive genetic value	0	$h^2\sigma^2_X$	
4) I_α, prediction of additive genetic value	0	$r^2_{TI}h^2\sigma^2_X$	*
5) $G_{A_\alpha} \mid I_\alpha = I_0$, additive genetic value for animals with index = I_0	I_0	$(1-r^2_{TI})h^2\sigma^2_X$	*
6) $G_{A_\alpha} - I_\alpha$, prediction error	0	$(1-r^2_{TI})h^2\sigma^2_X$	*

* r^2_{TI} depends on T, but for $T = G_{A_\alpha}$:

$$r^2_{TI} = \frac{\Sigma b_i \sigma_{X_i T}}{h^2 \sigma^2_X} \text{ if not inbred, and}$$

$$r^2_{TI} = \frac{\Sigma b_i \sigma_{X_i T}}{(1+F)h^2 \sigma^2_X} \text{ if inbred.}$$

CHAPTER 13

SELECTION WITH MORE THAN ONE TRAIT MEASURED

The contribution of genetic effects and environmental effects to the correlation between two traits can be described in the form of a simple model for phenotypic records of traits 1 and 2 expressed as differences from their means:

$$X_1 = P_1 - \mu_1 = \mu_1 + G_1 + E_1 - \mu_1 = G_1 + E_1$$

$$X_2 = P_2 - \mu_2 = \mu_2 + G_2 + E_2 - \mu_2 = G_2 + E_2 \text{ with}$$

$$\sigma^2_{P_1} = \sigma^2_{X_1} = \sigma^2_{G_1} + \sigma^2_{E_1}, \text{ which implies } \sigma_{G_1E_1} = 0;$$

$$\sigma^2_{P_2} = \sigma^2_{X_2} = \sigma^2_{G_2} + \sigma^2_{E_2}, \text{ which implies } \sigma_{G_2E_2} = 0; \text{ and}$$

$$\sigma_{P_1P_2} = \sigma_{X_1X_2} = \sigma_{G_1G_2} + \sigma_{E_1E_2}, \text{ which implies } \sigma_{G_1E_2} = \sigma_{G_2E_1} = 0.$$

Note that $\sigma_{G_1G_2}$ is the genetic covariance and $\sigma_{E_1E_2}$ is the environmental covariance between traits 1 and 2. In this chapter additive genetic effects will be assumed to be the only genetic effects. If other than additive genetic effects are present, the procedures described in this chapter can be changed easily to account for the other genetic effects.

The genetic correlation between traits 1 and 2 is:

$$r_g = \frac{\sigma_{G_1G_2}}{\sqrt{\sigma^2_{G_1}\,\sigma^2_{G_2}}}.$$

The environmental correlation is:

$$r_e = \frac{\sigma_{E_1E_2}}{\sqrt{\sigma^2_{E_1}\,\sigma^2_{E_2}}} \quad .$$

The phenotypic correlation is:

$$r_p = \frac{\sigma_{G_1G_2} + \sigma_{E_1E_2}}{\sqrt{(\sigma^2_{G_1} + \sigma^2_{E_1})(\sigma^2_{G_2} + \sigma^2_{E_2})}} = \frac{\sigma_{X_1X_2}}{\sqrt{\sigma^2_{X_1}\,\sigma^2_{X_2}}} \quad .$$

SELECTION FOR MORE THAN ONE TRAIT

There are several reasons for considering more than one trait in a selection program.

1. Records of other traits may be used in selecting for a single trait.
2. Several traits may be economically important so that joint selection is desirable.
3. Several economically important traits are to be improved but other traits are at an optimum level so that they should not be allowed to change.
4. In all cases, the correlated response in many traits may be of interest even if selection is not for all traits.

Definition of Overall Genetic Value and General Problem of Selection

If **m** traits have linear economic value, then overall or aggregate genetic value for animal α can be defined as:

$$T_\alpha = v_1G_{\alpha 1} + v_2G_{\alpha 2} + \cdots + v_mG_{\alpha m} = \sum_{j=1}^{m} v_j\,G_{\alpha j} \quad \text{where}$$

$G_{\alpha j}$ is the additive genetic value of animal α for trait **j**, and v_j is the net economic value per unit of trait **j**. As before, $\sigma^2_T = E[T^2]$.

Since the **G**'s are in units of measurement and the **v**'s are values per unit, the overall aggregate true or genetic value is in economic units -- for example, dollars or cents.

Next, suppose records are available for **N** traits measured on the animal to be evaluated [X_1, X_2, $\cdots$, X_N]. These records may, but need not, be included in the **m** traits included in overall economic value. The case when records on relatives are available will be discussed in Chapter 14. Thus, records on several traits are available to estimate T_α. The problem is, as before, to weight each record to estimate T_α with an index of the traits, i.e., $\hat{T}_\alpha = I = ß_1X_1 + ß_2X_2 + \cdots + ß_NX_N$, where the ß's are the weights which will maximize r_{TI} and ΔT. Several approaches to estimating **T** are equivalent, although proving the equivalence is not always easy. The general selection index procedure and properties as described in Chapter 7 apply to the multiple trait case as well as to the single trait case. The appendix to this chapter describes multiple trait selection index procedures in the notation of matrix algebra.

METHODS USING PHENOTYPIC RECORDS EXPRESSED AS DEVIATIONS FROM APPROPRIATE POPULATION AVERAGES

Records expressed as differences from population averages were considered for selection using records on relatives for only one trait. All traits measured on an animal will also be expressed as differences from their population averages.

Index Each Trait Separately

This method is perhaps the easiest to apply and to understand. The genetic value for each trait is estimated separately using all the traits with measurements, X_i ($i = 1, \cdots, N$).

Then the indexes, I_j, for the traits are substituted for genetic values, G_j, in the definition of T. For example, estimate G_j by $I_j = b_{j1}X_1 + b_{j2}X_2 + \cdots + b_{jN}X_N$. The first subscript on the b's refers to the trait being evaluated and the second subscript to the trait being weighted in the index. The equations which define the weights are the usual ones to maximize r_{TI} and to predict G_j :

$$\begin{array}{ccccccccc}
\sigma^2_{X_1} & b_{j1} + & \sigma_{X_1X_2} & b_{j2} + \cdots + & \sigma_{X_1X_N} & b_{jN} = & \sigma_{X_1G_j} \\
\sigma_{X_1X_2} & b_{j1} + & \sigma^2_{X_2} & b_{j2} + \cdots + & \sigma_{X_2X_N} & b_{jN} = & \sigma_{X_2G_j} \\
\vdots & & \vdots & & \vdots & & \vdots \\
\sigma_{X_1X_N} & b_{j1} + & \sigma_{X_2X_N} & b_{j2} + \cdots + & \sigma^2_{X_N} & b_{jN} = & \sigma_{X_NG_j}
\end{array}$$

Then, $I_j = b_{j1}X_1 + b_{j2}X_2 + \cdots + b_{jN}X_N$.

This procedure is repeated for all **m** traits with economic values so that:

$$\begin{aligned}
I_1 &= b_{11}X_1 + b_{12}X_2 + \cdots + b_{1N}X_N \\
I_2 &= b_{21}X_1 + b_{22}X_2 + \cdots + b_{2N}X_N , \\
&\vdots \\
I_m &= b_{m1}X_1 + b_{m2}X_2 + \cdots + b_{mN}X_N
\end{aligned}$$

Then because I_j estimates G_j, the I_j will be substituted for G_j in the economic equation;

$T = v_1G_1 + v_2G_2 + \cdots + v_mG_m$ so that the estimate of overall economic value

$\hat{T} = I = v_1I_1 + v_2I_2 + \cdots + v_mI_m$ where I is the overall index estimate of T.

In setting up the selection index equations to determine the weights for each index, the coefficients of the weights are the same ($\sigma^2_{X_1}$, $\sigma_{X_1X_2}$, etc.) no matter which trait is being indexed. The equations to find the weights change in the first subscript of the b's (the subscript for the trait being evaluated) and in the covariances on the RHS's. Depending on the trait to be evaluated, the RHS's correspond to the traits measured and are of the form:

$$\sigma_{X_iG_j} \quad (i=1, \cdots, N) \text{ for evaluation of trait } \mathbf{j}.$$

Because in this example X_i is measured on animal α and G_j is the genetic value for trait **j** of animal α and $X_i = G_i + E_i$, then $E(X_iG_j) = \sigma_{X_iG_j} = \sigma_{G_iG_j}$ which is the genetic covariance between traits **i** and **j** because $\sigma_{E_iG_j}$ is assumed to be zero. The covariance $\sigma_{G_iG_j}$ also can be written as $r_{g_{ij}} \sqrt{h_i^2 h_j^2}\, \sigma_{X_i}\, \sigma_{X_j}$ where σ_{X_i} and σ_{X_j} are the phenotypic standard deviations for traits **i** and **j**.

Although I can be calculated as $v_1I_1 + \cdots + v_mI_m$, obviously the overall index can be rewritten as $I = ß_1X_1 + ß_2X_2 + \cdots + ß_NX_N$ because each I_j contains all the X_i. In fact, appropriate multiplications and grouping of coefficients show:

$$ß_i = v_1b_{1i} + v_2b_{2i} + \cdots + v_mb_{mi}\text{, with, e.g.,}$$

$$ß_1 = v_1b_{11} + v_2b_{21} + \cdots + v_mb_{m1}\,.$$

The advantage of this method over the next one is that if economic values change, the equations to find the b's do not have to be solved again. The new economic values are simply substituted in the last step of the procedure, i.e., $I = v_1I_1 + \cdots + v_mI_m$.

If $v_j = 0$, there is no need to find the I_j to predict G_j since I_j will drop out of the overall index. The overall index, however, will include X_j if it is used in predicting genetic value of some other trait and, therefore, overall merit. If all $v_j = 0$ except for one trait, then overall economic value is defined as equal to the genetic value for the one trait that is being predicted by all the traits.

Index Overall Genetic Value Directly

The T in the general selection procedure to set up the index equations to find the appropriate weights for the X's is $T_\alpha = \sum_{j=1}^{m} v_j G_j$. The equations that define the ß's and maximize r_{TI} are as before:

$$\begin{aligned}
&\sigma^2_{X_1}\ \beta_1 + \sigma_{X_1X_2}\ \beta_2 + \cdots + \sigma_{X_1X_N}\ \beta_N = \sigma_{X_1T} \\
&\quad\vdots \qquad\qquad\qquad\qquad\qquad\quad \vdots \\
&\sigma_{X_1X_N}\ \beta_1 + \sigma_{X_2X_N}\ \beta_2 + \cdots + \sigma^2_{X_N}\ \beta_N = \sigma_{X_NT}
\end{aligned}$$

The coefficients of the weights on the LHS's are the same as when finding the weights to index each trait separately. The RHS's (the σ_{X_iT}, $i=1, \cdots, N$) are the covariances between X_i and the linear function, $v_1G_1 + \cdots + v_mG_m$. By the usual rules for finding the covariance between linear functions and by assuming no covariances between genetic values and environmental values:

$$\begin{aligned}
E(X_iT) = \sigma_{X_iT} &= \mathrm{Cov}(X_i, v_1G_1 + \cdots + v_mG_m) \\
&= v_1\sigma_{G_iG_1} + v_2\sigma_{G_iG_2} + \cdots + v_m\sigma_{G_iG_m} \\
&= \sum_{j=1}^{m} v_j\sigma_{G_iG_j}
\end{aligned}$$

Recall that $\sigma_{G_iG_j} = r_{g_{ij}}\sqrt{h_i^2h_j^2}\ \sigma_{X_i}\sigma_{X_j}$ and also note when $G_i = G_j$, that $\sigma_{G_iG_i} = \sigma^2_{G_i} = h_i^2\sigma^2_{X_i}$. Solving the equations for the ß's then gives the index $I = ß_1X_1 + ß_2X_2 + \cdots + ß_NX_N$ which is the same index as found earlier when indexing each trait separately and then weighting by economic value as $I = v_1I_1 + \cdots + v_mI_m$. Proof of the equivalence of the two procedures is in the appendix.

EXPECTED RESPONSE FROM SELECTION FOR ECONOMIC VALUE

Total response in economic value can be determined as before by either $\Delta T = D\sigma_I$ or with more difficulty with $\Delta T = r_{TI}D\sigma_T$ where σ_I^2 can be found as the variance of the linear function:

$$V(I) = E[(\sum_{i=1}^{N} \beta_i X_i)^2] \text{ and } r_{TI} = \sqrt{\Sigma \beta_i \sigma_{X_i T}/\sigma_T^2}$$

where the $\sigma_{X_i T}$ $(i=1, \cdots, N)$ are the covariances of linear functions and σ_T^2 is the variance of the linear function, $T = \sum_{j=1}^{m} v_j G_j$.

Often the expected correlated response for one or more traits is of interest when selecting for some overall defined economic value. For any index, whether the selection index or any other, the expected correlated genetic response for any trait **j** can be found by the regression of G_j on I:

$$\hat{G}_j = \mu_{G_j} + b_{G_j \cdot I}(I_{sel} - \mu_I) \quad ,$$

where the average selected I, I_{sel}, is:

$$\mu_I + D\sigma_I \text{ and } \hat{G}_j = \mu_{G_j} + \Delta G_j; \text{ thus, } \Delta G_j = [Cov(G_j,I)]D/\sigma_I \text{ where}$$

$$Cov(G_j, I) = Cov(G_j, \beta_1 X_1 + \cdots + \beta_N X_N) =$$

$$\beta_1 \sigma_{G_j G_1} + \cdots + \beta_2 \sigma_{G_j G_2} + \cdots + \beta_N \sigma_{G_j G_N}.$$

This formula holds for any trait whether included in T or I. However, the correlated responses of the traits included in T when weighted by their economic values will equal total economic response; i.e., $\Delta T = v_1 \Delta G_1 + v_2 \Delta G_2 + \cdots + v_m \Delta G_m$.

An example follows for selection for two traits. Included are examples of comparing expected correlated responses in the two traits when selection is for only one of them using either both traits or only one trait.

EXAMPLES

Selecting For More Than One Trait

Let milk yield = trait 1, type score = trait 2

$\sigma^2_{X_1} = (2000 \text{ lb})^2$ $\quad \sigma^2_{X_2} = (2\%)^2$ $\quad \sigma_{X_1X_2} = 400 \text{ lb } \%$ $\quad r_p = .1$

$\sigma^2_{G_1} = (1000 \text{ lb})^2$ $\quad \sigma^2_{G_2} = (1\%)^2$ $\quad \sigma_{G_1G_2} = 200 \text{ lb } \%$ $\quad r_g = .2$

$h^2_1 = .25$ $\quad h^2_2 = .25$

Suppose $v_1 = \$.025/\text{lb}$ and $v_2 = \$50./\%$.

Method 1. Find $I_1 = b_{11} X_1 + b_{12} X_2$

$$4{,}000{,}000\, b_{11} + 400\, b_{12} = \sigma_{X_1G_1} = \sigma^2_{G_1} = 1{,}000{,}000$$

$$400\, b_{11} + 4\, b_{12} = \sigma_{X_2G_1} = \sigma_{G_2G_1} = 200$$

Thus, $I_1 = .2475\, X_1 + 25.2525\, X_2$.

Then find $I_2 = b_{21} X_1 + b_{22} X_2$

$$4{,}000{,}000\, b_{21} + 400\, b_{22} = \sigma_{X_1G_2} = 200$$

$$400\, b_{21} + 4\, b_{22} = \sigma_{X_2G_2} = 1$$

Thus, $I_2 = .00002525\, X_1 + .2475\, X_2$ so that:

$$I = (.025)I_1 + (50.)I_2$$

$$= [.025(.2475) + 50(.00002525)]X_1 + [.025(25.2525) + 50(.2475)]X_2$$

$I = ß_1X_1 + ß_2X_2 = .00745\, X_1 + 13.006\, X_2$ will be the overall index.

Total Response

$\Delta G = D\, \sigma_I$; $\sigma^2_I = \beta^2_1 \sigma^2_{X_1} + \beta^2_2 \sigma^2_{X_2} + 2\beta_1\beta_2 \sigma_{X_1X_2} = 976$ and

$\Delta G = D \sqrt{976} = 31.24\, D$ ($), total expected economic response.

Correlated Responses

$$\Delta G_2 = \frac{[\mathrm{Cov}(G_2,I)]}{\sigma_I} D \quad \text{with}$$

$$\mathrm{Cov}(G_2,I) = \beta_1 \sigma_{G_2G_1} + \beta_2 \sigma^2_{G_2} = .00745(200) + 13.006(1) = 14.5;$$

$$\text{and } \Delta G_2 = \frac{14.5}{31.24} D = .464\% \text{ (D)} .$$

$$\Delta G_1 = \frac{[\mathrm{Cov}(G_1,I)]}{\sigma_I} D \quad \text{with } \mathrm{Cov}(G_1,I) = \beta_1 \sigma^2_{G_1} + \beta_2 \sigma_{G_1G_2} = 10051 ;$$

$$\text{and } \Delta G_1 = \frac{10051}{31.24} D = 321.7 \text{ lb (D)} .$$

As it should, $\Delta G = v_1 \Delta G_1 + v_2 \Delta G_2 = .025(321.7\ D) + 50.\ (.464\ D) =$

$$8.04\ D + 23.2\ D = 31.24\ D = D\sigma_I .$$

If the correlated response in another trait, e.g., fat test = trait 3, is of interest, then:

$$\Delta G_3 = \frac{[\mathrm{Cov}(G_3,I)]}{\sigma_I} D \quad \text{with } \mathrm{Cov}(G_3,I) = \beta_1 \sigma_{G_3G_1} + \beta_2 \sigma_{G_3G_2}.$$

If $r_{g_{13}} = -.6$, $r_{g_{23}} = .1$, $\sigma^2_{X_3} = (.3\%)^2$, and $h^2_3 = .5$;

then:

$$\sigma^2_{G_3} = .045,$$

$$\sigma_{G_1G_3} = r_{g_{13}} \sqrt{\sigma^2_{G_1} \sigma^2_{G_3}} = -.60 \sqrt{(1000)^2(.045)} = 127, \text{ and}$$

$$\sigma_{G_2G_3} = r_{g_{23}} \sqrt{\sigma^2_{G_2} \sigma^2_{G_3}} = .10 \sqrt{(1)^2(.045)} = .02121 .$$

Then: $\text{Cov}(G_3,I) = .00745(-127) + 13.006(.02121) = -.67$ and

$$\Delta G_3 = \frac{[(-.67)]}{(31.24} D = -.021\ D\ (\%).$$

Selecting For One Trait Using Two Traits

Suppose $v_2 = 0$, then v_1 can be any positive nonzero value because obviously v_1 will not change ranking based on $I = v_1 I_1$; unity is a convenient value for v_1 .

If $v_1 \neq 1$, then

$$\Delta G_1 = \frac{[\text{Cov}(G_1,I)]}{\sigma_I} D \quad \text{but } I = v_1 I_1 \text{ so that}$$

$$\Delta G_1 = \frac{[v_1 \text{Cov}(G_1,I_1)]}{v_1 \sigma_{I_1}} D = \frac{[\text{Cov}(G_1,I_1)]}{\sigma_{I_1}} D \quad \text{as for } v_1 = 1 .$$

Thus for $v_1 = 1$ and $v_2 = 0$: $I = I_1 = .2475\ X_1 + 25.2525\ X_2$.

Response

$$\Delta G_1 = \Delta I_1 = D\ \sigma_{I_1} \ ;$$

$$\sigma^2_{I_1} = (.2475)^2\sigma^2_{X_1} + (25.2525)^2\sigma^2_{X_2} + 2(.2475)(25.2525)\,\sigma_{X_1X_2} = 250{,}556 \text{ and}$$

$\sigma_{I_1} = 500.56$. Thus, $\Delta G_1 = 500.56\ D$ (lb)

Correlated Responses

When selecting for G_1 using X_1 and X_2:

$$\Delta G_2 = \frac{[\text{Cov}(G_2,I_1)]}{\sigma_{I_1}} D = \frac{[b_{11}\,\sigma_{G_2G_1} + b_{12}\,\sigma^2_{G_2}]}{\sigma_{I_1}} D = \frac{74.75}{500.56} D = .149\ D\ (\%)$$

$$\Delta G_3 = \frac{[\text{Cov}(G_3,I_1)]}{\sigma_{I_1}} D = \frac{[b_{11}\,\sigma_{G_3G_1} + b_{12}\,\sigma_{G_3G_2}]}{\sigma_{I_1}} D = \text{some practice ?}$$

Comparison With Selecting For One Trait Using Only Record of That Trait

Suppose $I_1 = b_1X_1 = h_1^2X_1 = .25\ X_1$

Response

$$\Delta G_1 = D\sigma_{I_1} \ ; \ \sigma_{I_1} = \sqrt{(.25)^2\sigma^2_{X_1}} = 500; \text{ thus,}$$

$$\Delta G_1 = 500\ D \text{ (lb)}$$

Correlated Response

$$\Delta G_2 = \frac{[\text{Cov}(G_2,I_1)]}{\sigma_{I_1}}D = \frac{[b_1\sigma_{G_2G_1}]}{\sigma_{I_1}}D = \frac{[.25(200)]}{500}D = .1\ D\ (\%)$$

This correlated response is the same response expected as when selecting indirectly for trait 2 using only trait 1 because the genetic covariance is positive.

Comparison With Selecting For One Trait Using A Record of Another Trait

Select for G_2 using X_1 by $I_2 = b_1X_1$:

$$\sigma^2_{X_1} b_1 = \sigma_{X_1G_2} = \sigma_{G_1G_2} \ ; b_1 = \frac{(\sigma_{G_1G_2})}{\sigma^2_{X_1}} = \frac{(200)}{4{,}000{,}000} = .00005$$

$$I_2 = .00005\ X_1 \ ; \ \Delta G_2 = D\sigma_{I_2} \ ; \ \sigma_{I_2} = \sqrt{(.00005)^2\sigma^2_{X_1}} = .1$$

Response

Thus, $\Delta G_2 = .1\ D\ (\%)$ as with selecting for G_1 using X_1 .

Summary

These examples illustrate the method of comparing different selection systems:

Selection for	Based on	ΔG_1	ΔG_2	ΔT for .025 = v_1 & 50 = v_2
.025 G_1 + 50 G_2	X_1, X_2	321.7 lb	.464 %	31.24 $
G_1	X_1, X_2	500.6 lb	.149 %	19.96 $
G_1	X_1	500.0 lb	.100 %	17.50 $
G_2	X_1	500.0 lb	.100 %	17.50 $
G_2	X_2	100.0 lb	.500 %	27.50 $

*All expected responses should be multiplied by D.

APPROXIMATE PROCEDURE FOR SELECTING FOR MORE THAN ONE TRAIT

Often the genetic correlations needed to find the weights for the index to estimate, $T = v_1G_1 + \cdots + v_mG_m$, are not known or are estimated with not much reliability. In addition, the equations to determine the weights may be difficult to solve if many traits are included in the index. An approximation which is easy to use is to index each trait using only records for that trait; then substitute those indexes into the economic value equation. This approximation can also be used when records of relatives are available as will be discussed later. The approximation is the same as the exact procedure when the phenotypic and genetic correlations among the traits are all zero. In fact, the assumption made to obtain the approximate index is that phenotypic and genetic covariances are very small.

When only one record is available on each trait of the animal to be evaluated, the indexes for each trait using only that trait are:

$$I_j = h_j^2 X_j.$$

The approximate overall index is:

$$I = v_1h_1^2X_1 + v_2h_2^2X_2 + \cdots + v_mh_m^2X_m .$$

The phenotypic records are weighted by the product of their economic values and heritabilities which would be the weights found by solving the equations for the ß's when all the phenotypic and genetic correlations are zero:

$$
\begin{aligned}
\sigma^2_{X_1}\beta_1 + \text{0's} \qquad\qquad &= v_1h_1^2\sigma^2_{X_1} + \text{0's} \\
\sigma^2_{X_2}\beta_2 + \text{0's} &= v_2h_2^2\sigma^2_{X_2} + \text{0's} \\
\vdots \qquad\qquad & \\
\text{0's} + \sigma^2_{X_m}\beta_m &= \text{0's} + v_mh_m^2\sigma^2_{X_m} .
\end{aligned}
$$

Some research has indicated that this approximation may be better than using poorly estimated genetic and phenotypic correlations to determine the weights for the "exact" procedure. Even if the correct genetic and phenotypic correlations are known, the approximate procedure may be nearly as good as the exact procedure and will be much easier to apply. In such cases, how good the approximation is may be found by calculating the correlated responses expected for each economic trait when selection is by the approximate method. The response in each trait can be compared to the response expected from the exact procedure. The responses for individual traits can be weighted by economic values to compare economic responses expected by the exact and approximate procedures.

Let $I_* = \beta_1^* X_1 + \cdots + \beta_m^* X_m$ be the approximate index. Then the correlated response for trait **j** using the approximate index is:

$$\Delta G_j = \frac{[\mathrm{Cov}(G_j, I_*)]}{\sigma_{I_*}} D \quad \text{where}$$

$$E(G_j I_*) = \mathrm{Cov}(G_j, I_*) = \beta_1^* \sigma_{G_j G_1} + \beta_2^* \sigma_{G_j G_2} + \cdots + \beta_m^* \sigma_{G_j G_m} .$$

$\sigma^2_{I_*}$ will be determined by the variance of a linear function. Care should be taken to include the correct phenotypic covariances such as $\sigma_{X_1 X_2}$ which were assumed to be zero in determining the approximate ß's.

Example of Approximate Procedure

Suppose that the selection is for milk and type score with variances and covariances as in the previous example. In the example, $v_1 = .025/\text{lb}$ and $v_2 = \$50./\%$.

The approximate procedure assumes the phenotypic and genetic covariances are zero.

Method 1:

The equations to find the index for milk are:

$$4{,}000{,}000\, b_{11}^* + 0\, b_{12}^* = 1{,}000{,}000$$

$$0\, b_{11}^* + 4\, b_{12}^* = 0 \quad ; \text{and}$$

$$I_{1_*} = .25\, X_1 + 0\, X_2 = h_1^2 X_1 .$$

The equations to find I_{2_*} , the index for type score, are:

$$4{,}000{,}000\, b_{21}^* + 0\, b_{12}^* = 0$$

$$0\, b_{21}^* + 4\, b_{22}^* = 1 \quad ; \text{and}$$

$$I_{2_*} = 0\, X_1 + .25\, X_2 = h_2^2 X_2 .$$

Then $I_* = v_1 I_{1_*} + v_2 I_{2_*} = v_1 h_1^2 X_1 + v_2 h_2^2 X_2 = .025(.25)\, X_1 + 50(.25)\, X_2$.

Note $I_* = .00625\, X_1 + 12.500\, X_2$ as compared to the optimum index of

$$I = .00745\, X_1 + 13.006\, X_2 .$$

Total and Correlated Responses

Total response would be computed incorrectly as:

$\Delta T = D\sigma_{I_*}$; because $\sigma^2_{I_*} = (.00625)^2\ \sigma^2_{X_1} + (12.5)^2\ \sigma^2_{X_2} + 0 = 781.25$ and $\sigma_{I_*} = 27.95$ are not correct because $\sigma_{X_1X_2}$ is not zero.

Actually $\sigma^2_{I_*} = (.00625)^2\ \sigma^2_{X_1} + (12.5)^2\ \sigma^2_{X_2} + 2(.00625)(12.5)\ \sigma_{X_1X_2} = 843.75$ with $\sigma_{I_*} = 29.05$.

The correct expected total response can be computed as:

$\Delta T = \dfrac{[\mathrm{Cov}(T,I_*)]}{\sigma_{I_*}} D$ where σ_{I_*} is computed using the correct $\sigma_{X_1X_2}$ and $T = v_1G_1 + v_2G_2$. The correct expected total response can also be computed from $v_1\,\Delta G_1 + v_2\,\Delta G_2$ where:

$$\Delta G_1 = \frac{[\mathrm{Cov}(G_1,I_*)]}{\sigma_{I_*}} D = \frac{[.00625\ \sigma^2_{G_1} + 12.5\ \sigma_{G_1G_2}]}{29.05} D = 301.2 \text{ lb (D) and}$$

$$\Delta G_2 = \frac{[\mathrm{Cov}(G_2,I_*)]}{\sigma_{I_*}} D = \frac{[.00625\ \sigma_{G_1G_2} + 12.5\ \sigma^2_{G_2}]}{29.05} D = .473\ \%\ (D)$$

Thus, expected $\Delta T = .025(301.2\ D) + 50(.473\ D) = 31.18\ D$. If the correlated responses are computed assuming $\sigma_{X_1X_2}$ and $\sigma_{G_1G_2} = 0$, the incorrect expected responses are:

$$\Delta G_1 = [\mathrm{Cov}(G_1, I_*)]D/\sigma_{I_*} = (.00625\ \sigma^2_{G_1})D/27.95 = 223.6 \text{ lb (D) and}$$

$$\Delta G_2 = [\mathrm{Cov}(G_2, I_*)]D/\sigma_{I_*} = (12.5\ \sigma^2_{G_2})D/27.95 = .447\%\ (D).$$

In the incorrect calculation, the genetic covariance term in the numerator was ignored and in the incorrect σ_{I_*}, the phenotypic covariance was ignored.

The three sets of calculated responses that were compared are summarized below:

1) using the correct covariances,
2) using zero covariances to approximate the index but using the correct covariances to compute response, and
3) using zero covariances when really not correct.

Computing of		Expected response / D		
index	response	ΔG_1 (lb)	ΔG_2 (%)	$\Delta T = v_1\Delta G_1 + v_2\Delta G_2$
correct	correct	321.7	.464	31.24
approximate	correct	301.2	.473	31.18
approximate	incorrect	223.6	.447	27.95

APPENDIX TO CHAPTER 13

MULTIPLE TRAIT INDEXES IN MATRIX FORM

Finally, a place where matrix algebra and matrix computations make problems of selection index much easier.

Let $\mathbf{g}$ = vector of genetic values of $\mathbf{m}$ economically important traits for animal α,

$\mathbf{x}$ = vector of phenotypic records on the same traits (in general, the vector could include records on different traits from those in $\mathbf{g}$; the algebra is a little more difficult so here $\mathbf{x}$ has same traits as $\mathbf{g}$) ,

$\mathbf{v}$ = vector of economic values for the traits in $\mathbf{g}$,

$\mathbf{G}$ = $(\mathbf{g}_1\ \mathbf{g}_2 \cdots \mathbf{g}_m)$, the genetic variance-covariance matrix with the $\mathbf{g}_i$, the columns of $\mathbf{G}$,

$\mathbf{G}$ = $E(\mathbf{gg}') = E[\mathbf{xg}']$ when $\mathbf{x}$ and $\mathbf{g}$ are the same traits, and

$\mathbf{P}$ = the phenotypic variance-covariance matrix of records in $\mathbf{x}$,

$\mathbf{P}$ = $E(\mathbf{xx}')$.

Define Overall Economic Values

$$T = \mathbf{v}'\mathbf{g} = \mathbf{g}'\mathbf{v}$$

Indirect Prediction of T

Predict T_j , the additive genetic value of trait $\mathbf{j}$;

The RHS's:

$$E[\mathbf{x}\ (\text{genetic value of trait } \mathbf{j})] = \mathbf{g}_j$$

Then:

$$\mathbf{P}\mathbf{b}_j = \mathbf{g}_j\,,\ \mathbf{b}_j = \mathbf{P}^{-1}\mathbf{g}_j,\ \text{and}\ I_j = \mathbf{b}'_j\mathbf{x}$$

Note that for traits $j = 1, \cdots, m$:

$$(\mathbf{b}_1\ \ \mathbf{b}_2 \cdots \mathbf{b}_m) = \mathbf{P}^{-1}\,(\mathbf{g}_1\ \ \mathbf{g}_2 \cdots \mathbf{g}_m)\ \text{and}$$

$$I_1 = \mathbf{b}'_1\mathbf{x}\,,\ I_2 = \mathbf{b}'_2\mathbf{x}\,,\ \cdots\,,\ I_m = \mathbf{b}'_m\mathbf{x}$$

Predict T from the I_j :

$$I = \mathbf{v}'\begin{pmatrix} I_1 \\ I_2 \\ \cdot \\ \cdot \\ \cdot \\ I_m \end{pmatrix} = \mathbf{v}'\begin{pmatrix} \mathbf{b}'_1 \\ \mathbf{b}'_2 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{b}'_m \end{pmatrix}\mathbf{x} = \mathbf{v}'\begin{pmatrix} \mathbf{g}'_1\mathbf{P}^{-1} \\ \mathbf{g}'_2\mathbf{P}^{-1} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{g}'_m\mathbf{P}^{-1} \end{pmatrix} = \mathbf{v}'\mathbf{G}\mathbf{P}^{-1}\mathbf{x}$$

Note from $\mathbf{v}'\begin{pmatrix} \mathbf{b}'_1 \\ \mathbf{b}'_2 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{b}'_m \end{pmatrix}\mathbf{x} = \boldsymbol{\beta}'\mathbf{x}$ that $\boldsymbol{\beta}' = \mathbf{v}'\begin{pmatrix} \mathbf{b}'_1 \\ \mathbf{b}'_2 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{b}'_m \end{pmatrix}$ and $\boldsymbol{\beta} = (\mathbf{b}_1\ \ \mathbf{b}_2 \cdots \mathbf{b}_m)\mathbf{v}$.

Direct Prediction of T

Predict $T = \mathbf{g'v}$;

RHS's :

$$E[xT] = E[\mathbf{xg'v}] = \mathbf{Gv}$$

Then :

$$\mathbf{P\beta} = \mathbf{Gv}\ ,\ \mathbf{\beta} = \mathbf{P^{-1}Gv}\ ,\ \text{and}\ \ I = \mathbf{\beta'x}$$

Note :

$$I = \mathbf{\beta'x} = \mathbf{v'GP^{-1}x}\ \text{ as with the indirect method.}$$

Calculations With I and T

$$V(I) = E[\mathbf{\beta'xx'\beta}] = \mathbf{\beta'}E[\mathbf{xx'}]\mathbf{\beta} = \mathbf{\beta'P\beta}$$

$$V(T) = E[\mathbf{v'gg'v}] = \mathbf{v'}E[\mathbf{gg'}]\mathbf{v} = \mathbf{v'Gv}$$

$$Cov(I,T) = E[\mathbf{\beta'xg'v}] = \mathbf{\beta'}E[\mathbf{xg'}]\mathbf{v} = \mathbf{\beta'Gv}$$

From $\mathbf{P\beta} = \mathbf{Gv}$, $V(I) = \mathbf{\beta'Gv} = Cov(I,T)$

From $\mathbf{Gv} = \mathbf{P\beta}$, $Cov(I,T) = \mathbf{\beta'P\beta} = V(I)$

$$r_{TI} = (\mathbf{\beta'Gv/v'Gv})^{.5}$$

CHAPTER 14

USING RECORDS ON ALL TRAITS OF RELATIVES

When all records of all traits are used to find the index for overall genetic value, covariances such as the one between a record for trait 1 of relative 1, P_{11}, and the record for trait 3 of relative 2, P_{23}, are needed to set up the equations to find the proper selection index weights. The usual models for such records are:

$P_{11} = G_{11} + E_{11}$ and $P_{23} = G_{23} + E_{23}$. Then,

$\text{Cov}(P_{11},P_{23}) = \text{Cov}(G_{11},G_{23}) + \text{Cov}(E_{11},E_{23}) + \text{Cov}(G_{11},E_{23}) + \text{Cov}(G_{23},E_{11})$.

All terms except $\text{Cov}(G_{11},G_{23})$ and, perhaps, $\text{Cov}(E_{11},E_{23})$ usually are assumed to be zero. The remaining covariance is the covariance between genetic value for trait 1 on relative 1 and genetic value for trait 3 on relative 2. If these are measured on the same animal, i.e., if relative 1 is relative 2 then the covariance is the additive genetic covariance between traits 1 and 2. In general, the additive genetic covariance between relatives is $a_{12}\sigma_{G_1 G_3}$, the product of the additive relationship between the relatives and the genetic covariance between the traits. This form of the covariance corresponds to the additive genetic covariance between relatives for the same trait, $a_{12}\sigma^2_G$. Thus, if only additive genetic effects are assumed, $\text{Cov}(G_{ij},G_{i'j'}) = a_{ii'}\sigma_{G_j G_{j'}}$ where i and i' are relatives i and i' and j and j' are traits j and j'. If the further assumption of no covariances among genetic and environmental effects and among environmental effects on different relatives is true

then,

$$\text{Cov}(P_{ij},P_{i'j'}) = a_{ii'}\,\sigma_{G_jG_{j'}} \quad \text{but when } i = i',$$

$$\text{Cov}(P_{ij},P_{ij'}) = \sigma_{G_jG_{j'}} + \sigma_{E_jE_{j'}} \quad \text{as before, and when } i = i' \text{ and } j = j',$$

$$\text{Cov}(P_{ij},P_{ij}) = \sigma^2_{P_j} = \sigma^2_{G_j} + \sigma^2_{E_j} \ .$$

The notation has been changed so that P_{ij} is a single phenotypic record for trait j on relative i since the selection index will use average records on all measured traits for all relative groups. Thus, $X_k = X_{ij}$ will be the average of records on relative group i for trait j (n_k records for each of p_k animals in the group). The overall index for $T = \sum_{i=1}^{m} v_i G_i$ will be $I = \sum_{k=1}^{N} ß_k X_k$. The equations which determine the ß's come as usual from maximizing r_{TI} or minimizing $E(T - I)^2$.

As in Chapter 13, finding the index directly and weighting the separate indexes for each of the economic traits by its economic value are equivalent. The procedure for finding the index for each trait separately using all the X's and then putting them together as $I = \sum_{i=1}^{m} v_i I_i$ will be described.

The basic step is to estimate $G_{\alpha j}$, the additive genetic value for trait j for animal α, from all X's (X_k, k=1, ..., N) as $I_{\alpha j} = b_{j1} X_1 + b_{j2} X_2 + \cdots + b_{jN} X_N$. The relationships among the relative groups and the animal being indexed must be known.

COEFFICIENTS OF THE SELECTION INDEX EQUATIONS

The general equations to find the b's to predict $G_{\alpha j}$ are:

$$\begin{aligned}
&\sigma^2_{X_1} b_{j1} + \sigma_{X_1X_2} b_{j2} + \cdots + \sigma_{X_1X_N} b_{jN} = \sigma_{X_1G_{\alpha j}} \\
&\quad\vdots \qquad\qquad\qquad\qquad\qquad\qquad\quad \vdots \\
&\sigma_{X_1X_N} b_{j1} + \sigma_{X_2X_N} b_{j2} + \cdots + \sigma^2_{X_N} b_{jN} = \sigma_{X_NG_{\alpha j}} \ .
\end{aligned}$$

If the usual simplifying assumptions are true, the variances and covariances can be written in terms of phenotypic and additive genetic variances of the traits and of the phenotypic and additive genetic covariances among the traits.

Variances of the X's, $\sigma^2_{X_k}$:

$$\sigma^2_{X_k} = \sigma^2_{X_{ij}} = \sigma^2_{P_j}\left(\frac{\frac{1+(n_k-1)r_j}{n_k} + (p_k-1)\, a_{ii}h^2_j}{p_k}\right) \text{ where}$$

$\sigma^2_{P_j}$ is the phenotypic variance for trait j,

r_j is the repeatability for trait j,

h^2_j is the heritability for trait j, and

a_{ii} is the additive relationship among animals in group i.

Covariances among the X's, $\sigma_{X_k X_{k'}}$:

There are three possible types of covariances:

1) If k = ij and k' = i'j (different relative groups, same trait j),

$\sigma_{X_k X_{k'}} = \sigma_{X_{ij} X_{i'j}} = a_{ii'}\sigma^2_{G_j}$ as before where $a_{ii'}$ is the additive relationship between groups i and i'.

2) If k = ij and k' = ij' (same group, different traits),

$$\sigma_{X_k X_{k'}} = \sigma_{X_{ij} X_{ij'}} = \frac{\sigma_{P_j P_{j'}} + (p_k-1)a_{ii}\sigma_{G_j G_{j'}}}{p_k} \text{ where}$$

$\sigma_{P_j P_{j'}}$ is the phenotypic covariance between traits j and j' and

a_{ii} is the relationship among animals in group i.

3) If k = ij and k' = i'j' (different groups, different traits),

$$\sigma_{X_k X_{k'}} = \sigma_{X_{ij} X_{i'j'}} = a_{ii'}\sigma_{G_j G_{j'}} \cdot$$

Covariances on the RHS's, $\sigma_{X_k G_{\alpha j}}$:

The covariances between the X_k and $G_{\alpha j}$ will be of two types:

1) If k = ij (same trait as $G_{\alpha j}$) then,

$$\sigma_{X_k G_{\alpha j}} = \sigma_{X_{ij} G_{\alpha j}} = a_{i\alpha} \sigma^2_{G_j}.$$

2) If k = ij' (different trait from G_j) then,

$$\sigma_{X_k G_{\alpha j}} = \sigma_{X_{ij'} G_{\alpha j}} = a_{i\alpha} \sigma_{G_{j'} G_j} \ .$$

Solving the equations gives the weights for $I_{\alpha j} = \sum_{k=1}^{N} b_{jk} X_k$, the index for animal α for trait j. This procedure is repeated for all economic traits. The coefficients of the b's are the same for all sets of equations but the RHS's change depending on the trait being indexed so that the b's are different.

Finally, $I_\alpha = \sum_{i=1}^{m} v_i \, I_{\alpha i} = \sum_{k=1}^{N} \beta_k \, X_k$.

EXPECTED RESPONSE FROM SELECTION

As usual $\Delta T = D \, \sigma_I$. Although σ^2_I is messy to compute, all the terms are found in the coefficients in the equations to find the weights; $\sigma^2_I = ß'Pß$.

The correlated response for any trait c can be computed as usual as

$$\Delta G_c = \frac{\mathrm{Cov}(G_{\alpha c}, I_\alpha)}{\sigma_I} D \ .$$

Again $\mathrm{Cov}(G_{\alpha c}, I_\alpha)$ is messy but can be computed as:

$$\mathrm{Cov}(G_{\alpha c}, I_\alpha) = ß_1 \mathrm{Cov}(G_{\alpha c}, X_1) + ß_2 \mathrm{Cov}(G_{\alpha c}, X_2) + \cdots + ß_N \mathrm{Cov}(G_{\alpha c}, X_N)$$

where $\mathrm{Cov}(G_{\alpha c}, X_k) = \mathrm{Cov}(G_{\alpha c}, X_{ij}) = a_{i\alpha} \sigma_{G_c G_j}$ and (if c = j) $= a_{i\alpha} \sigma^2_{G_c}$.

APPROXIMATE PROCEDURE WITH RELATIVES

As before, approximate weights can be determined easily by assuming the phenotypic and genetic covariances among the traits are zero. Then many of the equations to find the

weights have zero off-diagonal coefficients, i.e., all those between different traits. The RHS's are relationships times genetic variance or are zero if indexing each trait separately or are economic values times relationships times genetic variance if obtaining the overall index directly. The approximate procedure is the same as using records of relatives for only the trait being indexed. Then the approximate indexes for each trait (based only on records for that trait) are weighted by their economic values.

When the phenotypic and genetic covariances are really zero the approximate procedure is the same as the exact procedure. How much better the exact procedure is than the approximate procedure when the covariances are different from zero and are known can be determined by calculating the correlated responses by both procedures as was illustrated when only records on the animal were considered.

If the phenotypic and genetic covariances are estimated from a small amount of data so that they may be seriously in error, especially the genetic covariances, then the approximate procedure may be more accurate than using the exact procedure with incorrect covariances. The differences in the procedures, however, cannot be determined without knowing the correct covariances.

EXAMPLE OF APPROXIMATE PROCEDURE WITH RELATIVES

The following example with two traits measured on the animal and on 50 paternal half sibs (phs) will illustrate the exact and approximate procedures and will demonstrate how to compare the expected selection responses from both if the correct covariances are known.

Let:

X_1 = a record on trait 1 of animal α

X_2 = a record on trait 2 of animal α

X_3 = average of single records on trait 1 of 50 phs of α

X_4 = average of single records on trait 2 of the same 50 phs

Given: $\sigma^2_{P_1} = (2000 \text{ lb})^2$, $\sigma^2_{P_2} = (3\%)^2$, $h_1^2 = .25$, $h_2^2 = .36$

$r_g = .2$, $r_p = .1$, $v_1 = \$.05/\text{lb}$, $v_2 = \$25./\%$

Thus $\sigma_{P_1 P_2} = .1(2000)(3) = 600$, $\sigma_{G_1 G_2} = .2\sqrt{(.25)(.36)}(2000)(3) = 360$,

$\sigma^2_{G_1} = .25(2000)^2 = (1000)^2$, $\sigma^2_{G_2} = .36(3)^2 = (1.8)^2$.

Need: $I = ß_1X_1 + ß_2X_2 + ß_3X_3 + ß_4X_4$.

Exact Procedure

To find $I_1 = b_{11}X_1 + b_{12}X_2 + b_{13}X_3 + b_{14}X_4$ solve:

$$(2000)^2 b_{11} + 600\ b_{12} + .25(1000)^2 b_{13} + .25(360)\ b_{14} = (1)(1000)^2$$

$$600\ b_{11} + (3)^2\ b_{12} + .25(360)\ \ b_{13} + .25(1.8)^2 b_{14} = (1)(360);$$

$$.25(1000)^2 b_{11} + .25(360)\ b_{12} + \left(\frac{1+3.0625}{50}\right)(2000)^2 b_{13} + \left(\frac{600+4410}{50}\right) b_{14} = .25(1000)^2$$

$$.25(360)\ \ b_{11} + .25(1.8)^2 b_{12} + \left(\frac{600+4410}{50}\right) b_{13} + \left(\frac{1+4.41}{50}\right)(3)^2\ b_{14} = .25(360)$$

Then, $I_1 = .209\ X_1 + 20.56\ X_2 + .604\ X_3 - 6.22\ X_4$.

To find $I_2 = b_{21}X_1 + b_{22}X_2 + b_{23}X_3 + b_{24}X_4$, the equations have the same diagonal and off-diagonal coefficients but the RHS's become:

(1) (360), (1) $(1.8)^2$, .25 (360), and .25 $(1.8)^2$.

Then, $I_2 = .000032\ X_1 + .306\ X_2 - .0000096\ X_3 + .575\ X_4$

Thus, $I = .05\, I_1 + 25\, I_2$

$= \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$

$= .01125\, X_1 + 8.68\, X_2 + .02998\, X_3 + 14.07\, X_4$

$\sigma_I^2 = \beta' P \beta = \beta_1^2 (2000)^2 + \beta_2^2 (3)^2 + \cdots = 2314$ and $\sigma_I = 48.10$.

All of the terms, other than ß's in σ_I^2 came from the coefficients of the b's in the equations to find the b's, i.e., from P, the variance-covariance matrix of the X's.

Correlated Responses

$$\Delta T = \sigma_I D = 48.1 D .$$

$$\Delta G_1 = \frac{Cov(G_{\alpha 1}, I)}{\sigma_I} D$$

$$= \frac{[\beta_1 Cov(G_{\alpha 1}, X_1) + \beta_2 Cov(G_{\alpha 1}, X_2) + \beta_3 Cov(G_{\alpha 1}, X_3) + \beta_4 Cov(G_{\alpha 1}, X_4)]}{\sigma_I} D$$

$$= \frac{[\beta_1 (1)(1000)^2 + \beta_2 (1)(360) + \beta_3 (\frac{1}{4})(1000)^2 + \beta_4 (\frac{1}{4})(360)]}{\sigma_I} D$$

$$= \frac{23145}{48.10} D = 481.2 \text{ lb(D)} .$$

$$\Delta G_2 = \frac{Cov(G_{\alpha 2}, I)}{\sigma_I} D = .962 \%(D) .$$

The terms in $Cov(G_{\alpha 1}, I)$ other than ß's come from the RHS's of the equations to find I_1 as do terms in $Cov(G_{\alpha 2}, I)$.

Approximate Procedure

Pretend $r_g = r_p = 0$ so that $\sigma_{P_1 P_2} = \sigma_{G_1 G_2} = 0$.

The approximate procedure will be the correct procedure if $r_g = r_p = 0$.

To find $I_1^* = b_{11}^* X_1 + b_{12}^* X_2 + b_{13}^* X_3 + b_{14}^* X_4$ solve:

$$(2000)^2 b_{11}^* + 0\, b_{12}^* + .25(1000)^2 b_{13}^* + 0\, b_{14}^* = (1)(1000)^2$$

$$0\, b_{11}^* + (3)^2 b_{12}^* + 0\, b_{13}^* + .25(1.8)^2 b_{14}^* = 0;$$

$$.25(1000)^2 b_{11}^* + 0\, b_{12}^* + .08125\,(2000)^2 b_{13}^* + 0\, b_{14}^* = .25(1000)^2$$

$$0\, b_{11}^* + .25(1.8)^2 b_{12}^* + 0\, b_{13}^* + .1082\,(3)^2 b_{14}^* = 0\,.$$

Notice that:

$I_1^* = .212\, X_1 + .606\, X_3$ uses only records on trait 1.

To find I_2^* the RHS's change to 0, $(1)(1.8)^2$, 0, and $(1/4)(1.8)^2$. Then,

$I_2^* = .308\, X_2 + .575\, X_4$ uses only records on trait 2.

Then, $I^* = .05\, I_1^* + 25\, I_2^* = .0106\, X_1 + 7.71\, X_2 + .0303\, X_3 + 14.39\, X_4$

Note that $ß_1^* = v_1\, b_{11}^*$, $ß_2^* = v_2\, b_{22}^*$, $ß_3^* = v_1\, b_{13}^*$ and $ß_4^* = v_2\, b_{24}^*$.

If $r_g = r_p = 0$ then $\sigma_{I*}^2 = \sum_{i=1}^{4} (ß_i^*)^2 \sigma_{X_i}^2 = 1824.5$ and $\sigma_{I*} = 42.71$. Then,

$$\Delta G_1 = \frac{COV(G_{\alpha 1}, I*)}{\sigma_{I*}} D = \frac{ß_1^* (1)(1000)^2 + ß_3^* .25(1000)^2}{42.71} D = 425.7 \text{ lb (D)}$$

$$\Delta G_2 = \frac{COV(G_{\alpha 2}, I*)}{\sigma_{I*}} D = \frac{ß_2^* (1)(1.8)^2 + ß_4^* (\frac{1}{4})(1.8)^2}{42.71} D = .857\ \% \text{ (D)}$$

If r_g or $r_p \neq 0$, then σ^2_{I*} and the correlated responses as computed above are incorrect because the genetic and phenotypic covariances have been ignored. The correct correlated responses and correct σ^2_{I*} can be computed if the correct covariances are known even if the index is not the best index:

$$\sigma^2_{I*} = \sum_{i=1}^{4} (ß_i^*)^2 \sigma^2_{X_i} + 2 \sum_{i>j} \sum ß_i^* ß_j^* \sigma_{X_i X_j}$$ [The correct covariance terms in the variance of the linear function I* have been added].

$\sigma^2_{I*} = 1824.5 + \underline{595.3} = 2419.8$ and $\sigma_{I*} = 49.19$.

$$COV(G_{\alpha 1}, I^*) = ß_1^* (1)(1000)^2 + ß_2^* (1)(360) + ß_3^* (.25)(1000)^2 + ß_4^* (.25)(360)$$
$$= 18181.8 + \underline{4070.1} = 22252.5 \quad \text{and}$$

$$COV(G_{\alpha 2}, I^*) = ß_1^* (1)(360) + ß_2^* (1)(1.8)^2 + ß_3^* (.25)(360) + ß_4^* (25)(1.8)^2$$
$$= 36.62 + \underline{6.54} = 43.16 .$$

These covariances now contain the genetic covariances between the two traits. Thus, the correct correlated responses are:

$$\Delta G_1 = \frac{22252.5}{49.19} D = 452.4 \text{ lb (D)} \quad \text{and} \quad \Delta G_2 = \frac{43.16}{49.19} D = .877 \text{ \% (D)} .$$

The three sets of calculated responses can be compared and are summarized below.

Computing of index	response	ΔG_1 (lb)	ΔG_2 (%)	$\Delta T = .05\ \Delta G_1 + 25\ \Delta G_2$ ($)
Correct	correct	481.2	.962	48.10
Approximate	correct	452.4	.877	44.55
Approximate	incorrect	425.7	.857	42.71

METHODS USING STANDARDIZED RECORDS

All the procedures that have been described can be applied using standardized variables. Much of the early technical literature expressed records of all traits in a standardized form. In addition, the amount of relative economic emphasis on each trait expressed in value per standard deviation of the trait, v_i', is appropriate when the records are standardized. As will be seen the final selection index is the same whether records are standardized or not.

If records on the traits are standardized the equations to find the selection index weights can be written in terms of r_p's, r_g's, and h^2's. Standardizing puts all variables on the same scale with mean, zero, and variance, 1. The standardized records are then expressed as fractions of standard deviations above or below the mean.

Records on traits 1 and 2, say X_1 and X_2 are standardized as follows where Y_1 and Y_2 are the standardized records for traits 1 and 2:

$$Y_1 = \frac{X_1 - \mu_{X_1}}{\sigma_{X_1}} \quad \text{and} \quad Y_2 = \frac{X_2 - \mu_{X_2}}{\sigma_{X_2}}.$$

If E[ ] is the expected or average value of what is in the parenthesis then:

$$E(Y_1) = \mu_{Y_1} = E\,\frac{(X_1 - \mu_{X_1})}{\sigma_{X_1}} = \frac{1}{\sigma_{X_1}}\left[E(X_1) - E(\mu_{X_1})\right] = \frac{1}{\sigma_{X_1}}\left[\mu_{X_1} - \mu_{X_1}\right] = 0$$

Similarly, $\mu_{Y_2} = 0$.

By definition $\sigma_Y^2 = E(Y - \mu_Y)^2 = E\left(\frac{X - \mu_X}{\sigma_X}\right)^2 = \frac{1}{\sigma_X^2} E(X - \mu_X)^2$. But by definition

$E(X - \mu_X)^2 = \sigma_X^2$ so that $\sigma_Y^2 = E(Y - \mu_Y)^2 = \sigma_X^2/\sigma_X^2 = 1$.

Correlations And Genetic Variances Of Standardized Variables

Standardized records can be used for estimating overall genetic value or genetic value for a particular trait. If v_i' is the economic value per phenotypic standard deviation of trait i then:

$$T = v_1' \frac{G_1}{\sigma_{X_1}} + v_2' \frac{G_2}{\sigma_{X_2}} + \cdots \quad \text{and}$$

$$\hat{T} = I' = \beta_1' Y_1 + \beta_2' Y_2 + \cdots \quad \text{where } Y_i = \frac{X_i}{\sigma_{X_i}} .$$

The equations to find the weights (β_i', i=1, ..., N) are:

$$\sigma^2_{Y_1} \beta_1' + \sigma_{Y_1 Y_2} \beta_2' + \cdots = \sigma_{Y_1 T}$$

$$\sigma_{Y_1 Y_2} \beta_1' + \sigma^2_{Y_2} \beta_2' + \cdots = \sigma_{Y_2 T} .$$

$$\vdots \qquad\qquad \vdots$$

The diagonal coefficients, $\sigma^2_{Y_i}$, for single records are all 1.

The phenotypic covariances, $\sigma_{Y_i Y_j} = r_{p_{ij}}$:

$$\sigma_{Y_i Y_j} = \text{Cov}\left(\frac{X_i}{\sigma_{X_i}}, \frac{X_j}{\sigma_{X_j}}\right) = \frac{1}{\sigma_{X_i}\sigma_{X_j}} \sigma_{X_i X_j} = r_{p_{ij}} .$$

The covariances on the RHS's, $\sigma_{Y_i T}$, will be made up of functions of genetic covariances and often a genetic variance of standardized variables, i.e.,

$$\sigma_{Y_i T} = v_1' \text{Cov}\left(\frac{G_1}{\sigma_{X_1}}, \frac{G_i}{\sigma_{X_i}}\right) + v_2' \text{Cov}\left(\frac{G_2}{\sigma_{X_2}}, \frac{G_i}{\sigma_{X_i}}\right) + \cdots + v_m' \text{Cov}\left(\frac{G_m}{\sigma_{X_m}}, \frac{G_i}{\sigma_{X_i}}\right)$$

The genetic covariance between standardized variables i and j is:

$$\text{Cov}\left(\frac{G_i}{\sigma_{X_i}},\frac{G_j}{\sigma_{X_j}}\right) = \frac{1}{\sigma_{X_i}\sigma_{X_j}}\,\sigma_{G_iG_j} \quad .$$

Multiplying both numerator and denominator by $\sigma_{G_i}\sigma_{G_j}$ gives:

$$\text{Cov}\left(\frac{G_i}{\sigma_{X_i}},\frac{G_j}{\sigma_{X_j}}\right) = r_{g_{ij}}\sqrt{h_i^2 h_j^2} \quad .$$

When i=j, the equation shows that the genetic variance of a standardized variable is h_i^2. Thus, the equations to find ($ß_i'$, i=1, ..., N) can be rewritten in terms of genetic and phenotypic correlations and heritabilities.

The genetic and environmental variances for a standardized variable, Y, can be shown in another way.

$$Y = \frac{X-\mu_X}{\sigma_X} = \frac{G_X + E_X}{\sigma_X} = \frac{G_X}{\sigma_X} + \frac{E_X}{\sigma_X} \text{ so that } \frac{G_X}{\sigma_X} = G_Y \text{ and } \frac{E_X}{\sigma_X} = E_Y \, .$$

$$\sigma_Y^2 = \frac{\sigma_{G_X}^2}{\sigma_X^2} + \frac{\sigma_{E_X}^2}{\sigma_X^2} \text{ where } \frac{\sigma_{G_X}^2}{\sigma_X^2} = \sigma_{G_Y}^2 = h^2 \text{ and } \frac{\sigma_{E_X}^2}{\sigma_X^2} = \sigma_{E_Y}^2 = 1 - h^2$$

Equivalences In Using Standardized And Un-standardized Records

The value of a standard deviation of trait j is v_j' and the value per unit of trait j is v_j which gives the equivalence, $v_j' = v_j\,\sigma_{X_j}$.

The indexes I' and I both estimate T and so are equal, i.e.,

$$I' = ß_1'Y_1 + \cdots + ß_N'Y_N = I = ß_1X_1 + \cdots + ß_NX_N \, .$$

Thus $\beta_j' Y_j = \beta_j X_j$ but $Y_j = X_j/\sigma_{X_j}$ so that $\beta_j' = \beta_j \sigma_{X_j}$ and $\beta_j = \beta_j'/\sigma_{X_j}$.

If each standardized trait is indexed separately as:

$$\hat{G}_j/\sigma_{X_j} = I_j' = b_{j1}' Y_1 + \cdots + b_{jN}' Y_N ;$$

then, I_j' is in fractions of phenotypic standard deviations so that:

$$I_j' \sigma_{X_j} = I_j = \hat{G}_j \text{ and } I_j' = I_j/\sigma_{X_j} .$$

To estimate overall economic value:

$\hat{T} = I' = v_1' I_1' + \cdots + v_m' I_m'$ and $\hat{T} = I = v_1 I_1 + \cdots + v_m I_m$, so that $v_j I_j = v_j' I_j'$.

Because $I_j' \sigma_{X_j} = I_j$ and $v_j' = v_j \sigma_{X_j}$; then $v_j' I_j' = v_j \sigma_{X_j} I_j/\sigma_{X_j} = v_j I_j$.

When selection is for one trait using several traits, an economic weight of one is often assigned to the trait, say j, being evaluated. If standardized records are used, then $v_j' = 1 = v_j \sigma_{X_j}$ with $v_j = 1/\sigma_{X_j}$. However, what is wanted in the nonstandardized index is $v_j = 1$. To convert the standardized index with $v_j' = 1$ to a non-standardized index with $v_j = 1$, the standardized index must be multiplied by σ_{X_j} which would be equivalent to $v_j' = \sigma_{X_j}$. Then, as stated,

$I_j = \sigma_{X_j} I_j' = \sigma_{X_j} [(b_{j1}' X_1/\sigma_{X_1}) + (b_{j2}' X_2/\sigma_{X_2}) + \cdots + (b_{jN}' X_N/\sigma_{X_N})]$ so that

$b_{j1} = b_{j1}' \sigma_{X_j}/\sigma_{X_1}$; $b_{j2} = b_{j2}' \sigma_{X_j}/\sigma_{X_2}$; $\cdots$; $b_{jN} = b_{jN}' \sigma_{X_j}/\sigma_{X_N}$.

Equations To Find Weights

As before, the overall index can be found either by indexing each trait or by indexing T directly. The equations to find the weights to index each trait separately are given below assuming one standardized record on each trait of the animal being evaluated.

To find the weights for $I_j' = b_{j1}' Y_1 + \cdots + b_{jN}' Y_N$ solve:

$$b_{j1}' + r_{p12} b_{j2}' + \cdots + r_{p1N} b_{jN}' = r_{gj1} \sqrt{h_1^2 h_j^2}$$

$$\begin{aligned}
r_{p12} b_{j1}' + \quad & b_{j2}' + \cdots + r_{p2N} b_{jN}' = r_{gj2} \sqrt{h_2^2 h_j^2} \\
& \vdots \qquad\qquad\qquad \vdots \\
r_{p1N} b_{j1}' + r_{p2N} b_{j2}' + \cdots + \quad & b_{jN}' = r_{gjN} \sqrt{h_N^2 h_j^2} \quad .
\end{aligned}$$

Repeat for each trait (j=1, ..., m). Then the overall index:

$$I' = I = v_1' I_1' + v_2' I_2' + \cdots + v_m' I_m' \quad .$$

Note also that:

$$I' = I = \beta_1' Y_1 + \beta_2' Y_2 + \cdots + \beta_N' Y_N = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_N X_N \, .$$

If solving directly for I' then the equations are:

$$\begin{aligned}
\beta_1' + r_{p12} \beta_2' + \cdots + r_{p1N} \beta_N' &= \sigma_{Y_1 T} \\
\vdots \qquad\qquad \vdots \qquad & \vdots \\
r_{p1N} \beta_1' + r_{p2N} \beta_2' + \cdots + \beta_N' &= \sigma_{Y_N T}
\end{aligned}$$

where the typical RHS:

$$\sigma_{Y_i T} = v_1' r_{gi1} \sqrt{h_1^2 h_i^2} + v_2' r_{gi2} \sqrt{h_2^2 h_i^2} + \cdots + v_m' r_{gim} \sqrt{h_m^2 h_i^2} \quad .$$

Correlated Responses

Standardized correlated responses can be computed in terms of standardized covariances and then converted to the usual units by multiplication by the standard

deviation. For example, for trait c the standardized correlated response is:

$$\Delta G'_c = \frac{\mathrm{Cov}(G'_c, I')}{\sigma_{I'}} D = \left(\frac{\beta'_1 r_{g_{c1}} \sqrt{h_1^2 h_c^2} + \cdots + \beta'_N r_{g_{cN}} \sqrt{h_N^2 h_c^2}}{\sigma_{I'}} \right) D .$$

Then $\Delta G_c = \sigma_{X_c} \Delta G'_c$.

If unstandardized genetic covariances are used and because the β_i can be determined:

$$\Delta G_c = \frac{\mathrm{Cov}(G_c, I)}{\sigma_I} D = \frac{\beta_1 \sigma_{G_c G_1} + \cdots + \beta_N \sigma_{G_c G_N}}{\sigma_I} D \quad \text{as before.}$$

Note $\sigma_I = \sigma_{I'}$ since $I = I'$.

The following example illustrates use of standardized variables when three traits are measured on the animal being evaluated and when the three traits have nonzero economic values. A part of the example also illustrates the consequences of assuming the genetic and phenotypic correlations are zero.

Example

Given: $v'_1 = 3, \quad v'_2 = 2, \quad v'_3 = 1$ (the relative economic values of phenotypic standard deviations of the traits)

$\sigma_{X_1} = 6, \quad \sigma_{X_2} = 5, \quad \sigma_{X_3} = 4, \quad h_1^2 = .7, \quad h_2^2 = .8, \quad h_3^2 = .9$

$r_{p_{12}} = .1, \; r_{p_{13}} = .2, \; r_{p_{23}} = .3, \; r_{g_{12}} = .6, \; r_{g_{13}} = .5, \; r_{g_{23}} = .4$

Find: $I' = v'_1 I'_1 + v'_2 I'_2 + v'_3 I'_3 = \beta'_1 Y_1 + \beta'_2 Y_2 + \beta'_3 Y_3$.

For I_1' solve: $b_{11}' + .1\, b_{12}' + .2\, b_{13}' = .7$

$$.1\, b_{11}' + b_{12}' + .3\, b_{13}' = .6\sqrt{(.7)(.8)}$$

$$.2\, b_{11}' + .3\, b_{12}' + b_{13}' = .5\sqrt{(.7)(.9)} \quad \text{, so that}$$

$$I_1' = .632\, Y_1 + .335\, Y_2 + .170\, Y_3 \ .$$

For I_2' the RHS's become $.6\sqrt{(.8)(.7)}$, .8, and $.4\sqrt{(.8)(.9)}$ with:

$$I_2' = .366\, Y_1 + .751\, Y_2 + .041\, Y_3 \ .$$

For I_3' the RHS's are $.5\sqrt{(.9)(.7)}$, $.4\sqrt{(.9)(.8)}$, and .9 with:

$$I_3' = .223\, Y_1 + .066\, Y_2 + .835\, Y_3 \ .$$

To find I' directly solve:

$$ß_1' + .1\, ß_2' + .2\, ß_3' = 3(.7) + 2(.6)\sqrt{(.7)(.8)} + (1)(.5)\sqrt{(.7)(.9)}$$

$$.1\, ß_1' + ß_2' + .3\, ß_3' = 3(.6)\sqrt{(.8)(.7)} + 2(.8) + (1)(.4)\sqrt{(.8)(.9)}$$

$$.2\, ß_1' + .3\, ß_2' + ß_3' = 3(.5)\sqrt{(.9)(.7)} + 2(.4)\sqrt{(.9)(.8)} + (1)(.9)$$

Then $I' = 2.85\, Y_1 + 2.57\, Y_2 + 1.43\, Y_3 = I$ and $\sigma_{I'} = 4.7$.

The correlated response in trait 2 is:

$$\Delta G_2' = \frac{(2.85)(.6)\sqrt{(.8)(.7)} + (2.57)(.8) + (1.43)(.4)\sqrt{(.8)(.9)}}{4.7} D$$

$$= .813\, D$$

and $\Delta G_2 = 5(.813)D = 4.065\, D$.

Now assume r_g's and r_p's = 0 as for the approximate index.

The overall equations reduce to $ß_1' = 3(.7)$, $ß_2' = 2(.8)$, and $ß_3' = (1)(.9)$

so that $I' = 2.1\, Y_1 + 1.6\, Y_2 + .9\, Y_3$, $\sigma_{I'}^2 = 7.78$ and $\sigma_{I'} = 2.79$,

$\text{Cov}(G_2', I') = 1.28$ and $\Delta G_2 = 2.29\, D$, if r_g's and r_p's are really zero.

If r_g's and r_p's are not zero but are as given above and if $I' = 2.1\ Y_1 + 1.6\ Y_2 + .9\ Y_3$ then $\sigma^2_{I'} = 7.78 + \underline{2.29}$, $\sigma_{I'} = 3.17$; $\text{Cov}(G_2, I') = 2.53$ and $\Delta G_2 = (5)\frac{2.53}{3.17}\ D = 3.99\ D$ as compared to 4.06 D using the best index and to the 2.29 D calculated by using zero correlations when calculating expected correlated responses.

ANOTHER STANDARDIZATION

Some research papers have used another standardization procedure which gives all of the standardized variables a genetic variance of 1 and a mean value of zero. The standardization is to subtract the mean and divide by the genetic standard deviation:

$$Y = \frac{X - \mu_X}{\sigma_{G_X}} = \frac{G_X}{\sigma_{G_X}} + \frac{E_X}{\sigma_{G_X}} \quad .$$

$$V\left(\frac{G_X}{\sigma_{G_X}}\right) = 1 \quad \text{and} \quad V\left(\frac{E_X}{\sigma_{G_X}}\right) = \frac{1-h^2}{h^2} \quad \text{so that} \quad V(Y) = \frac{1}{h^2} \quad .$$

For standardized records on two traits, the phenotypic covariance is:

$$\text{Cov}(Y_1, Y_2) = \text{Cov}\left(\frac{X_1}{\sigma_{G_{X_1}}}, \frac{X_2}{\sigma_{G_{X_2}}}\right) = \frac{\sigma_{X_1 X_2}}{\sigma_{G_{X_1}} \sigma_{G_{X_2}}} = r_{P12} \frac{1}{\sqrt{h_1^2 h_2^2}} \quad .$$

$$\text{Genetic covariance is } \text{Cov}(G_{Y_1}, G_{Y_2}) = \text{Cov}\left(\frac{G_{X_1}}{\sigma_{G_{X_1}}}, \frac{G_{X_2}}{\sigma_{G_{X_2}}}\right) = \frac{\sigma_{G_{X_1} G_{X_2}}}{\sigma_{G_{X_1}} \sigma_{G_{X_2}}} = r_{g12}.$$

If records are standardized in this way, the equations that determine the correct weights (b''_{ji}, i=1, ..., N) when selecting for trait j using standardized records on the animal to be evaluated, $I''_j = b''_{j1} Y_1 + \cdots + b''_{jN} Y_N$, are:

$$\frac{1}{h_1^2} b''_{j1} + \frac{1}{\sqrt{h_1^2 h_2^2}} r_{p12} b''_{j2} + \cdots + \frac{1}{\sqrt{h_1^2 h_N^2}} r_{p1N} b''_{jN} = r_{gj1}$$

$$\vdots \qquad\qquad\qquad\qquad\qquad\qquad \vdots \quad \vdots$$

$$\frac{1}{\sqrt{h_1^2 h_N^2}} r_{p1N} b''_{j1} + \frac{1}{\sqrt{h_2^2 h_N^2}} r_{p2N} b''_{j2} + \cdots + \frac{1}{h_N^2} b''_{jN} = r_{gjN}$$

Note that when j = i, $r_{gji} = 1$.

The extension of this procedure to $T = \sum_j v''_j G_j$ is straightforward. Each economic value is given in terms of value per genetic standard deviation, v''_j . The index for trait j in standardized form can be converted back to non-standardized form as:

$I_j = \sigma_{G_j} I''_j$. Similarly $\hat{T} = I = v_1 I_1 + \cdots + v_m I_m$.

CHAPTER 15

SELECTION INDEX FOR CATEGORICAL DATA

Some traits are subjectively scored on an either-or basis; that is, they are assigned to a discrete category. For example, calving difficulty for a particular birth might be scored in one of three categories: 1, no difficulty; 2, some difficulty; or 3, great difficulty.

One method of analysis is to simply assign a single score to each birth. Two simple ways of doing that have been used: a) the score is the same as the category, e.g., a some difficulty birth would be scored as a "2" and b) the categories are assigned economic values and the score is the economic value associated with the category, e.g., if category 2 has economic value -$20, then the score for a some difficulty birth would be -20. Note in case a) that a linear scale of 1, 2, 3 for economic value is implied. In both cases the usual selection index procedure can be used if the appropriate heritability is known.

A better procedure, however, with the selection index is to consider each category being considered as a separate trait scored as zero or one. There will, however, be automatic covariances among the categories. Categorical data have a multinomial distribution. If there are only two categories the distribution is binomial.

VARIANCES AND COVARIANCES

The phenotypic variances of, and covariances among, the categories are determined by the probabilities of being scored in each category. These probabilities are also the population means when each category is scored as a zero (the attribute is absent) or as a one (the attribute is present). Suppose the fractions in each category (means) are π_1, π_2, and π_3 (the Greek symbol pi is used here to denote proportion). Then the phenotypic variances and covariances are:

$$\begin{pmatrix} \sigma^2_{y_1} & \sigma_{y_1 y_2} & \sigma_{y_1 y_3} \\ \sigma_{y_1 y_2} & \sigma^2_{y_2} & \sigma_{y_2 y_3} \\ \sigma_{y_1 y_3} & \sigma_{y_2 y_3} & \sigma^2_{y_3} \end{pmatrix} = \begin{pmatrix} \pi_1(1-\pi_1) & -\pi_1\pi_2 & -\pi_1\pi_3 \\ -\pi_1\pi_2 & \pi_2(1-\pi_2) & -\pi_2\pi_3 \\ -\pi_1\pi_3 & -\pi_2\pi_3 & \pi_3(1-\pi_3) \end{pmatrix} .$$

The sum of the variance and covariances in any row (or column) is zero because $\pi_1 + \pi_2 + \pi_3 = 1$.

The genetic variances and covariances follow the same pattern although they are not determined by the means:

$$\begin{pmatrix} \sigma^2_{g_1} & \sigma_{g_1 g_2} & \sigma_{g_1 g_3} \\ \sigma_{g_1 g_2} & \sigma^2_{g_2} & \sigma_{g_2 g_3} \\ \sigma_{g_1 g_3} & \sigma_{g_2 g_3} & \sigma^2_{g_3} \end{pmatrix} .$$

The sum of the variances and covariances in any row (or column) is zero. Such a property results in what is known as a lack of independence and such variance-covariance matrices are singular. The practical result is that instead of using all the traits to predict the value

for any one trait as is usual for evaluation using multiple traits, all traits (categories) except one are used as will be illustrated.

SELECTION INDEX TO PREDICT CATEGORY FREQUENCIES

The selection indexes predict differences from the means as follows:

$$I_1 = \hat{g}_1 - \pi_1 = b_{11}(X_1 - \pi_1) + b_{12}(X_2 - \pi_2) + b_{13}(X_3 - \pi_3)$$

$$I_2 = \hat{g}_2 - \pi_2 = b_{21}(X_1 - \pi_1) + b_{22}(X_2 - \pi_2) + b_{23}(X_3 - \pi_3)$$

$$I_3 = \hat{g}_3 - \pi_3 = b_{31}(X_1 - \pi_1) + b_{32}(X_2 - \pi_2) + b_{33}(X_3 - \pi_3) \quad .$$

The probabilities can be predicted by adding the means to the indexes as follows:

$$\hat{g}_1 = I_1 + \pi_1$$

$$\hat{g}_2 = I_2 + \pi_2$$

$$\hat{g}_3 = I_3 + \pi_3$$

Note that $\hat{g}_1 + \hat{g}_2 + \hat{g}_3 = 1$, and $I_1 + I_2 + I_3 = 0$. The multiple trait observation (X_1, X_2, X_3) is:

(1, 0, 0) if scored in category 1

(0, 1, 0) if scored in category 2, and

(0, 0, 1) if scored in category 3.

The same properties hold, for example, in the case of sire evaluation from **p** half sib progeny except that X_1, X_2, and X_3 are the fractions of progeny scored in categories 1, 2, and 3.

Then, $$\sigma^2_{X_i} = \frac{\sigma^2_{y_i} + (p-1)\left(.25\,\sigma^2_{g_i}\right)}{p} \quad \text{and}$$

$$\sigma_{X_i X_j} = \frac{\sigma_{y_i y_j} + (p-1)\left(.25\,\sigma_{g_i g_j}\right)}{p} \quad .$$

Again $\Sigma X_i = 1$, and $\sum_{j=1}^{3} \sigma_{X_i X_j} = 0$ for all rows (or columns). The RHS's, $\sigma_{X_i T}$, are determined as usual as the additive relationship times the appropriate column of the genetic variance-covariance matrix.

Because of the lack of independence one less equation than number of categories is used. The weight (e.g., b_{i3}) corresponding to the equation that is left out is set equal to zero.

EXAMPLE

An example may help clarify the procedure. Suppose for some trait with three categories that $\pi_1 = .5$, $\pi_2 = .3$, and $\pi_3 = .2$. Thus, the phenotypic variances and covariances are:

$$\begin{pmatrix} .25 & -.15 & -.10 \\ -.15 & .21 & -.06 \\ -.10 & -.06 & .16 \end{pmatrix}.$$

Assume the genetic variances and covariances are:

$$\begin{pmatrix} .05 & -.03 & -.02 \\ -.03 & .07 & -.04 \\ -.02 & -.04 & .06 \end{pmatrix}.$$

When the equation for trait 3 is set equal to zero, the selection index equations to determine the weights are:

	RHS's for g_1	g_2	g_3
$.25b_1 - .15b_2 =$	.05	-.03	-.02
	;	;	
$-.15b_1 + .21b_2 =$	-.03	.07	-.04

The indexes are:

$$\text{For } g_1: \quad I_1 = .20(X_1 - \pi_1) + 0(X_2 - \pi_2) + 0(X_3 - \pi_3)$$

$$\text{For } g_2: \quad I_2 = .14(X_1 - \pi_1) + .433(X_2 - \pi_2) + 0(X_3 - \pi_3)$$

$$\text{For } g_3: \quad I_3 = -.34(X_1 - \pi_1) - .433(X_2 - \pi_2) + 0(X_3 - \pi_3)$$

and $\hat{g}_1 = I_1 + \pi_1$, $\hat{g}_2 = I_2 + \pi_2$, $\hat{g}_3 = I_3 + \pi_3$.

For an animal scored in category 1; $X_1 = 1$, $X_2 = 0$, and $X_3 = 0$ so that:

$$\hat{g}_1 = .20(1 - .5) + 0(0 - .3) + 0(0 - .2) + .5 = .60,$$

$$\hat{g}_2 = .14(1 - .5) + .433(0 - .3) + 0(0 - .2) + .3 = .24, \text{ and}$$

$$\hat{g}_3 = -.34(1 - .5) - .433(0 - .3) + 0(0 - .2) + .2 = .16 .$$

For an animal scored in category 2; $X_1 = 0$, $X_2 = 1$, and $X_3 = 0$ so that:

$$\hat{g}_1 = .20(0 - .5) + 0(1 - .3) + 0(0 - .2) + .5 = .40,$$

$$\hat{g}_2 = .14(0 - .5) + .433(1 - .3) + 0(0 - .2) + .3 = .5331, \text{ and}$$

$$\hat{g}_3 = -.34(0 - .5) - .433(1 - .3) + 0(0 - .2) + .2 = .0669 .$$

For an animal scored in category 3; $X_1 = 0$, $X_2 = 0$, and $X_3 = 1$ so that:

$$\hat{g}_1 = .20(0 - .5) + 0(0 - .3) + 0(1 - .2) + .5 = .40,$$

$$\hat{g}_2 = .14(0 - .5) + .433(0 - .3) + 0(1 - .2) + .3 = .10, \text{ and}$$

$$\hat{g}_3 = -.34(0 - .5) - .433(0 - .3) + 0(1 - .2) + .2 = .50 .$$

If any other two equations had been used, e.g., X_2 and X_3 with $b_1 = 0$, rather than $b_3 = 0$, the evaluation would have been exactly the same. The appropriate RHS's to predict g_1, g_2, and g_3 would have been:

$$\begin{pmatrix} -.03 & .07 & -.04 \\ & ; \quad ; & \\ -.02 & -.04 & .06 \end{pmatrix} .$$

The procedure for finding the weights for sire evaluation would be similar. The RHS's would be divided by one-half and the LHS's computed as indicated earlier.

PREDICTION OF PROGENY FREQUENCIES

Prediction of progeny frequencies from a particular mating would be the same as averaging the evaluations of the sire and dam:

$$\text{Fraction in category 1} = \frac{\hat{g}_{1S} + \hat{g}_{1D}}{2}$$

$$\text{Fraction in category 2} = \frac{\hat{g}_{2S} + \hat{g}_{2D}}{2}$$

$$\text{Fraction in category 3} = \frac{\hat{g}_{3S} + \hat{g}_{3D}}{2}$$

Economic weights for any of the three cases, animal, sire, or progeny, can be assigned after the frequencies have been predicted. Suppose $v_1 = 60$, $v_2 = -5$, and $v_3 = -100$. Then, in the previous example, for an animal scored in category 2, (0, 1, 0), the aggregate economic value is predicted to be: 60(.40) - 5(.5331) - 100(.0669) = -45.5655.

CHAPTER 16

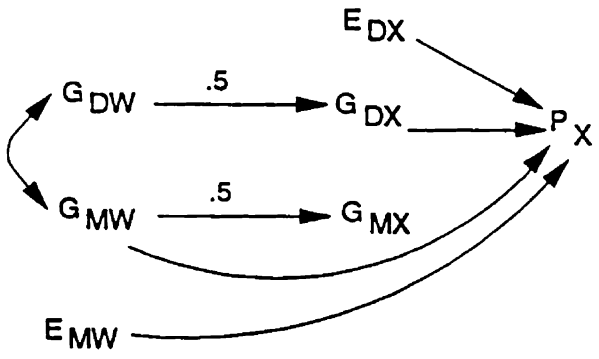
SELECTION FOR EMBEDDED TRAITS: MATERNAL EFFECTS

Some traits cannot be measured directly. An example is a maternal trait which makes up part of the environmental effects on the record of an offspring. Such traits are embedded traits. Selection for embedded traits, however, can be accomplished with the selection index. The procedure appears to be somewhat of a hybrid between single trait and multiple trait selection. The general selection index procedure can certainly be applied. Four examples of embedded traits will be discussed: in this chapter, the maternal effects model and in the following chapters, the grandmaternal effects model, the fetal effects model, and the cytoplasmic effects model.

SELECTION WHEN TRAITS ARE INFLUENCED BY MATERNAL EFFECTS

The maternal effect of the mother often has an effect on the phenotype of the offspring. This effect is genetic with respect to the mother but acts as an environmental effect on the offspring. This effect of the mother is in addition to the genetic effect of the sample one-half of her genes transmitted to her offspring. In turn, part of the maternal effect may be genetic and part may be environmental. See Willham (1963) for a complete development.

The following diagram illustrates the various effects when W is the dam of X.



P_X is the phenotype of animal X, E_{DX} is the non-maternally caused environmental effect, G_{DX} is the direct genetic effect associated with the genotype of X, G_{MW} is the genetic maternal effect of W, E_{MW} is the environmental effect on maternal ability of W, G_{MX} is the genetic maternal ability of X which is not measured, and G_{DW} is the direct genetic effect associated with W, the dam of X. Note that $G_{DX} = G_X$ and $G_{MW} + E_{MW} + E_{DX} = E_X$ of the usual model, $P_X = G_X + E_X$.

The direct trait, D, and the indirect maternal trait, M, can be considered to be two traits which may be correlated. The genetic value for trait M is measured one generation later than the direct effect D and is embedded in the phenotypic measurement of the animal that carries the direct effect.

GENETIC COVARIANCES BETWEEN RELATIVES WITH MATERNAL EFFECTS

Let the model for a record on animal X be,

$$P_X = G_{DX} + E_{DX} + G_{MW} + E_{MW}$$

and the model for a record on animal Y be,

$$P_Y = G_{DY} + E_{DY} + G_{MZ} + E_{MZ}$$

where animal Z is the mother of Y and W is the mother of X.

The rules for the covariance of linear functions provide the genetic covariance between P_X and P_Y. If all environmental covariances are zero:

$$COV(P_X, P_Y) = COV(G_{DX}, G_{DY}) + COV(G_{DX}, G_{MZ}) + COV(G_{MW}, G_{DY}) + COV(G_{MW}, G_{MZ}).$$

These terms involve covariances between genetic effects for the same trait on relatives and covariances between genetic effects for different traits on relatives.

In terms of genetic variance and covariance components:

$$COV(G_{DX}, G_{DY}) = a_{XY}\sigma^2_{D10} + a^2_{XY}\sigma^2_{D20} + d_{XY}\sigma^2_{D01} + \cdots,$$

$$COV(G_{MW}, G_{MZ}) = a_{WZ}\sigma^2_{M10} + a^2_{WZ}\sigma^2_{M20} + d_{WZ}\sigma^2_{M01} + \cdots,$$

$$COV(G_{DX}, G_{MZ}) = a_{XZ}\sigma_{DM10} + a^2_{XZ}\sigma_{DM20} + d_{XZ}\sigma_{DM01} + \cdots, \text{ and}$$

$$COV(G_{MW}, G_{DY}) = a_{WY}\sigma_{DM10} + a^2_{WY}\sigma_{DM20} + d_{WY}\sigma_{DM01} + \cdots.$$

The a's and d's are the usual additive and dominance relationships. The genetic variances are labelled with the trait, e.g., σ^2_{D10} is the additive genetic variance of the direct trait, D. The genetic covariances are labelled with both traits, e.g., σ_{DM10} is the covariance between the additive genetic effects for trait D and trait M.

If only additive genetic effects are considered, a simpler notation will be:

$$COV(P_X,P_Y) = a_{XY}\sigma^2_{D10} + a_{WZ}\sigma^2_{M10} + (a_{XZ} + a_{WY})\sigma_{DM10}$$

$$= a_{XY}\sigma^2_{G_D} + a_{WZ}\sigma^2_{G_M} + (a_{XZ} + a_{WY})\sigma_{G_DG_M}.$$

EXAMPLES OF COVARIANCES BETWEEN RELATIVES

Animal With Itself

This covariance is a variance that contains the genetic variance plus environmental variance.

Note: X = Y, W = Z, and $a_{XY} = 1$, $a_{WZ} = 1$, $a_{XW} = .5$, and $a_{YZ} = .5$ so that:

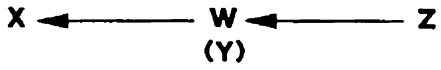
$$COV(P_X,P_X) = \sigma^2_{G_D} + \sigma^2_{G_M} + (.5 + .5)\sigma_{G_DG_M} + \sigma^2_{E_M} + \sigma^2_{E_D} = \sigma^2_P.$$

In terms of the usual P = G + E model:

$$\sigma^2_G = \sigma^2_{G_D} \text{ and } \sigma^2_E = \sigma^2_{G_M} + \sigma_{G_DG_M} + \sigma^2_{E_M} + \sigma^2_{E_D}.$$

Dam-progeny Covariance Considering Only Additive Genetic Effects

X is the progeny, W is the dam, Z is the dam's dam, Y is also the dam of X:



The genetic parts of the models for P_X and P_Y are:

$$P_X = \underset{(MY)}{G_{MW}} + G_{DX} \text{ and } P_Y = \underset{(W)}{G_{DY}} + \underset{(DW)}{G_{MZ}}.$$

$$\underset{(XW)}{a_{XY}} = .5, \quad \underset{(YZ)}{a_{WZ}} = .5, \quad a_{XZ} = .25, \quad \underset{(WW)}{a_{WY}} = 1.$$

Then:

$$COV(P_X,P_Y) = .5\sigma^2_{G_D} + .5\sigma^2_{G_M} + (.25 + 1)\sigma_{G_DG_M}.$$

The covariance between relatives may contain a genetic covariance between the direct and maternal traits. This covariance can be negative and, if so, would mask part of the additive genetic variances for the direct and maternal traits.

The additive genetic correlation between D and M is:

$$r_{g_{D,M}} = \frac{\sigma_{G_D G_M}}{\sqrt{\sigma^2_{G_D}\,\sigma^2_{G_M}}} .$$

Since the maximum absolute value of r_g is 1:

$$\sqrt{\sigma^2_{G_D}\,\sigma^2_{G_M}} \geq |\sigma_{G_D G_M}| .$$

Thus, a negative estimate of the offspring-dam covariance is possible if the negative value of $\sigma_{G_D G_M}$ is large enough.

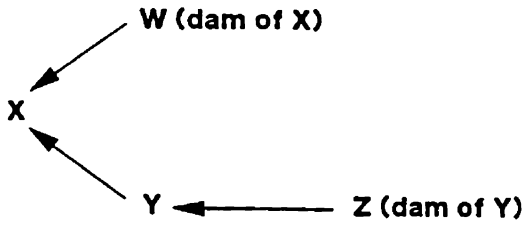
If maternal effects are not zero, the usual procedure of doubling the offspring on dam regression to estimate heritability can give a biased estimate of heritability of the direct trait; i.e.,

$$h^2 = \frac{2\left[.5\,\sigma^2_{G_D} + .5\,\sigma^2_{G_M} + 1.25\,\sigma_{G_D G_M}\right]}{\sigma^2_P} = \frac{\sigma^2_{G_D}}{\sigma^2_P} + \frac{\sigma^2_{G_M} + 2.5\,\sigma_{G_D G_M}}{\sigma^2_P}$$

There may also be other possible genetic causes for bias in this estimate due to higher order genetic variances, such as σ^2_{20} .

Sire-progeny Covariance

Now X is progeny of dam W and sire Y which has dam Z:



Because $a_{XY} = .5$, $a_{WZ} = 0$, $a_{XZ} = .25$, $a_{YW} = 0$,

then $Cov(P_X,P_Y) = .5\sigma^2_{G_D} + .25\sigma_{G_DG_M}$ which is quite different from the offspring-dam covariance.

PRACTICE PROBLEMS FOR COVARIANCES WITH MATERNAL EFFECTS

The following problems illustrate some concepts of covariance among relatives when maternal traits are important.

Problems

1. Estimate $\sigma^2_{G_D}$, $\sigma^2_{G_M}$ and $\sigma_{G_DG_M}$

   From: Covariance between paternal half sibs = 20

   Covariance between full sibs = 30

   Covariance between offspring and sire = 30

2. Given: $\sigma^2_{G_D} = 80$, $\sigma^2_{G_M} = 40$, $\sigma_{G_D G_M} = -20$

Show all calculations in computing the covariances between:

a) offspring and dam

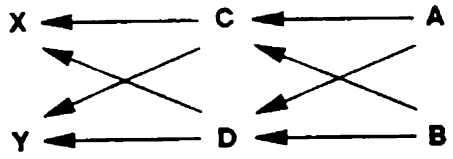
b) offspring and sire

c) full sibs

d) maternal half sibs

e) paternal half sibs

f) X and Y in diagram



Solutions

1. Cov (phs) $= .25\sigma^2_{G_D} + 0\sigma^2_{G_M} + 0\sigma_{G_D G_M} = 20$ [1]

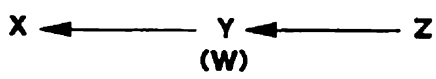
Cov (full sibs) $= .5\,\sigma^2_{G_D} + 1\sigma^2_{G_M} + 1\sigma_{G_D G_M} = 30$ [2]

Cov (offs, sire) $= .5\,\sigma^2_{G_D} + 0\sigma^2_{G_M} + .25\sigma_{G_D G_M} = 30$ [3]

From [phs]: $\sigma^2_{G_D} = 4(20) = 80$

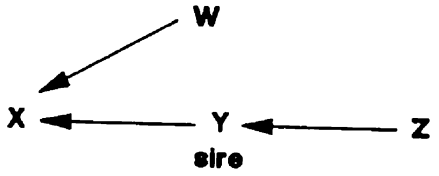
From [phs] and [offs, sire]: $.5(80) + .25\sigma_{G_D G_M} = 30$; $\sigma_{G_D G_M} = -40$

From [full sibs]: $.5\,(80) + \sigma^2_{G_M} + (-40) = 30$; $\sigma^2_{G_M} = 30$

2. a) 

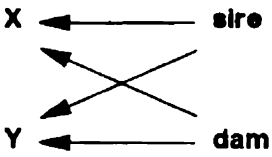
$a_{XY} = .5 \quad a_{WZ} = .5 \quad a_{XZ} = .25 \quad a_{YW} = 1$

Cov (offspring-dam) = .5(80) + .5(40) + (.25 + 1) (-20) = 35

b) 

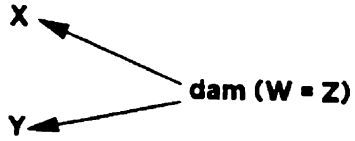
$a_{XY} = .5 \quad a_{WZ} = 0 \quad a_{XZ} = .25 \quad a_{YW} = 0$

Cov (offspring-sire) = .5(80) + 0 + (.25 + 0) (-20) = 35

c) 

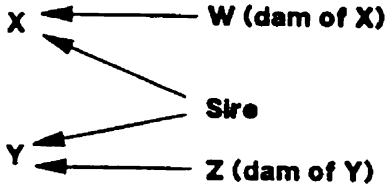
$a_{XY} = .5 \quad a_{WZ} = 1 \quad a_{XZ} = .5 \quad a_{YW} = .5$

Cov (fullsibs) = .5(80) + (1)(40) + (.5 + .5) (-20) = 60

d) 

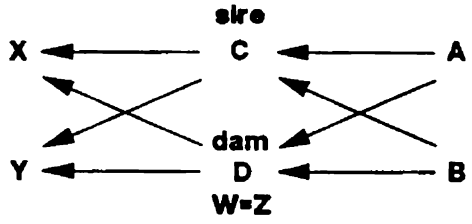
$a_{XY} = .25 \quad a_{WZ} = 1 \quad a_{XZ} = .5 \quad a_{YW} = .5$

Cov (mat. half sibs) = .25(80) + (1) (40) + (.5 + .5) (-20) = 40

e) 

$a_{XY} = .25 \quad a_{WZ} = 0 \quad a_{XZ} = 0 \quad a_{YW} = 0$

Cov (pat. half sibs) = .25(80) + 0 + (0 + 0) = 20

f) 

$a_{XY} = .75 \quad a_{WZ} = 1 \quad a_{XZ} = .75 \quad a_{YW} = .75$

Cov (X and Y) = .75(80) + 1(40) + (.75 + .75)(-20) = 70

The following table gives the additive relationships that are coefficients of $\sigma^2_{G_D}$, $\sigma^2_{G_M}$, and $\sigma_{G_D G_M}$ in the covariances between the pairs of relatives in the previous problem under the assumption that only additive genetic effects contribute to direct and maternal genetic effects.

CONTRIBUTION OF DIRECT AND MATERNAL ADDITIVE GENETIC VARIANCE AND COVARIANCE TO THE COVARIANCE BETWEEN RELATIVES

Genetic $\mathbf{Cov(P_X,P_Y) = a_{XY}\sigma^2_{G_D} + a_{WZ}\sigma^2_{G_M} + (a_{XZ} + a_{YW})\,\sigma_{G_D G_M}}$

P_X,P_Y	a_{XY}	a_{WZ}	a_{XZ}	a_{YW}
P_X,P_X (with self)	1	1	.50	.50
Progeny, dam	.50	.50	.25	1
Progeny, sire	.50	0	.25	0
Full sibs	.50	1	.50	.50
Maternal sibs	.25	1	.50	.50
Paternal sibs	.25	0	0	0

SELECTION FOR THE DIRECT AND MATERNAL TRAITS

Selection For The Direct Trait

The records used for the selection index will correspond to X_i, the average of single records of p_i animals in relative group i.

$$\text{RHS's: } \sigma_{X_i G_{D\alpha}} = a_{i\alpha}\sigma^2_{G_D} + a_{W_i\alpha}\sigma_{G_D G_M}$$

where W_i is the dam of i and α is the animal being evaluated.

Diagonal Coefficients

The variance of an average will be expanded by the maternal variance and maternal-direct covariance:

$$\sigma^2_{X_i} = \{\sigma^2_P + (p_i - 1)[a_{ii'}\sigma^2_{G_D} + a_{W_iW_{i'}}\sigma^2_{G_M} + (a_{iW_{i'}} + a_{i'W_i})\,\sigma_{G_DG_M}]\}/p;$$

where $a_{ii'}$ is additive relationship among members of the group, $a_{W_iW_{i'}}$ is additive relationship among dams of the group, and $a_{iW_{i'}}$ is relationship of animal in group to another's dam.

Off-diagonal Coefficients

The off-diagonal coefficients will be expanded similarly:

$$\sigma_{X_iX_j} = a_{ij}\sigma^2_{G_D} + a_{W_iW_j}\sigma^2_{G_M} + (a_{iW_j} + a_{jW_i})\sigma_{G_DG_M}.$$

Selection For The Maternal Trait

RHS's: $\sigma_{X_iG_{M\alpha}} = a_{i\alpha}\sigma_{G_DG_M} + a_{W_i\alpha}\sigma^2_{G_M}$

Diagonal and off-diagonal coefficients will be as in selection for the direct trait.

Correlated Responses When Selecting For G_D Or G_M

If selection is for G_{D_α} and $X_i (i = 1, \ldots, N)$ is the average of relative group i and the relationship of dam of relative i to α is $a_{W_i\alpha}$ then the response in G_D will be $\Delta G_D = \sigma_{I_D}\, D$ (the selection differential D is different from the subscript D which refers to the direct trait), where $I_D = b_{1D}X_1 + \cdots + b_{ND}X_N$.

The response in G_D can also be computed as:

$$\Delta G_D = \frac{\text{Cov}\left(G_{D\alpha}, I_D\right)}{\sigma_{I_D}} D = \frac{\sum_{i=1}^{N} b_{iD}\ \text{Cov}\left(G_{D\alpha}, X_i\right)}{\sigma_{I_D}} D$$

where $\text{Cov}\left(G_{D\alpha}, X_i\right) = a_{i\alpha}\sigma^2_{G_D} + a_{W_i\alpha}\sigma_{G_D G_M}$.

Similarly the correlated response in G_M can be predicted as:

$$\Delta G_M = \frac{\text{Cov}\left(G_{M\alpha}, I_D\right)}{\sigma_{I_D}} D = \frac{\sum_{i=1}^{N} b_{iD}\ \text{Cov}\left(G_{M\alpha}, X_i\right)}{\sigma_{I_D}} D$$

where $\text{Cov}\left(G_{M\alpha}, X_i\right) = a_{i\alpha}\sigma_{G_D G_M} + a_{W_i\alpha}\sigma^2_{G_M}$.

If selection is for $G_{M\alpha}$ by $I_M = \sum_{i=1}^{N} b_{iM} X_i$, then :

$$\Delta G_D = \frac{\text{Cov}\left(G_{D\alpha}, I_M\right)}{\sigma_{I_M}} D = \frac{\sum_{i=1}^{N} b_{iM}\ \text{Cov}\left(G_{D\alpha}, X_i\right)}{\sigma_{I_M}} D \text{ and}$$

$$\Delta G_M = \sigma_{I_M} D \qquad = \frac{\sum_{i=1}^{N} b_{iM}\ \text{Cov}\left(G_{M\alpha}, X_i\right)}{\sigma_{I_M}} D \ .$$

The following examples illustrate computations for these concepts and also show how to compute the effect of bias in heritability estimates if maternal effects are ignored.

PROBLEMS OF SELECTION FOR DIRECT GENETIC EFFECTS

Assume: $\sigma^2_{G_D} = 80$, $\sigma^2_{G_M} = 40$, $\sigma_{G_D G_M} = 40$, $\sigma^2_P = 500$

Also assume heritability is estimated in the usual way as twice the regression of offspring record on dam's record.

Problems

1. a) Use this biased estimate of heritability (genetic variance) to find the usual weights for indexing additive genetic value ignoring maternal effects from the animal's own record, X_1, and the sire's record, X_2.
   b) What is the bias in the calculation of expected response due to selection by the usual but now biased procedure of calculating genetic gain?
   c) Use the incorrect index found in 1a) but the correct variances and covariances to find the expected correlated responses in G_D and G_M.
2. a) Use the correct variances and covariances as given to find weights for indexing direct genetic value (G_D) from X_1 and X_2.
   b) Use the correct index for G_D and the correct variances and covariances to find the expected correlated responses in G_D and G_M.
3. Repeat 1) and 2) when $\sigma_{G_D G_M} = -40$.

Solutions

Heritability is incorrectly estimated from twice regression of offspring on dam record:

$$h^2_* = 2\,\text{Cov}(\text{offspring, dam})/\sigma^2_P = 2[.5(80) + .5(40) + 1.25(40)]/500 = .44$$

1. a) $b_1 + .5(.44)\, b_2 = .44$

$.22\, b_1 + \quad b_2 = .5(.44)$

$I = .4115\, X_1 + .1295\, X_2 = \hat{G}_\alpha$, $\sigma_I = 10.24$ is the apparent standard deviation of index.

The actual σ_I is 9.92 since $\sigma_{X_1 X_2} = 50$ while $h^2_* = .44$ implies $\sigma_{X_1 X_2} = 110$.

b) $\Delta G = 10.24\, D$ would be the usual prediction based on $h^2_* = .44$.

c) $$\Delta G_D = \frac{b_1 \mathrm{Cov}(G_{D\alpha}, X_1) + b_2 \mathrm{Cov}(G_{D\alpha}, X_2)}{\sigma_I} D =$$

$$\frac{.4115[(1)(80) + (.5)(40)] + .1295[(.5)(80) + (.25)(40)]}{9.92} D = 4.80\, D$$

$$\Delta G_M = \frac{b_1 \mathrm{Cov}(G_{M\alpha}, X_1) + b_2 \mathrm{Cov}(G_{M\alpha}, X_2)}{\sigma_I} D =$$

$$\frac{.4115[(1)(40) + (.5)(40)] + .1295[(.5)(40) + (.25)(40)]}{9.92} D = 2.88\, D$$

2. a) $500\, b_1 + [.5(80) + (.25)(40)]\, b_2 = (1)\, 80 + (.5)(40)$

$50\, b_1 + \quad 500\, b_2 = .5\,(80) + (.25)(40)$

$I = .1919\, X_1 + .0808\, X_2 = \hat{G}_{D\alpha}$ and $\sigma_I = 4.82$

b) $$\Delta G_D = \frac{b_1 \mathrm{Cov}(G_{D\alpha}, X_1) + b_2 \mathrm{Cov}(G_{D\alpha}, X_2)}{\sigma_I} D$$

$$= \frac{.1919(100) + .0808(50)}{4.82} D = 4.82\, D$$

$$\Delta G_M = \frac{b_1 \mathrm{Cov}(G_{M\alpha}, X_1) + b_2 \mathrm{Cov}(G_{M\alpha}, X_2)}{\sigma_I} D$$

$$= \frac{.1919(60) + .0808(30)}{4.82} D = 2.89\, D$$

3. $\sigma^2_{G_D} = 80,\ \sigma^2_{G_M} = 40,\ \sigma_{G_D G_M} = -40,\ \sigma^2_P = 500$

Now the biased heritability from twice offspring on parent regression is:

$$h^2_* = 2\ \text{Cov(offspring, dam)}/\sigma^2_P = 2[.5(80) + .5(40) + 1.25(-40)]/500 = .04$$

(1.a)
$$b_1 + .5(.04)\, b_2 = .04$$
$$.02\, b_1 + \quad b_2 = .5(.04)$$

$I = .0396\, X_1 + .0192\, X_2 = \hat{G}_\alpha$;

$\sigma_I = .9918$ is apparent σ_I, but the correct $\sigma_I = 1.007$.

(1.b) $\Delta G = .9918\, D$ is the usual prediction with $h^2_* = .04$

(1.c) The correct expected responses are:

$$\Delta G_D = \frac{b_1 \text{Cov}(G_{D\alpha}, X_1) + b_2 \text{Cov}(G_{D\alpha}, X_2)}{\sigma_I} D =$$

$$\frac{.0396[(1)(80) - .5(40)] + .0192[.5(80) - .25(40)]}{1.007} D = 2.931\, D$$

$$\Delta G_M = \frac{b_1 \text{Cov}(G_{M\alpha}, X_1) + b_2 \text{Cov}(G_{M\alpha}, X_2)}{\sigma_I} D =$$

$$\frac{.0396[(1)(-40) + .5(40)] + .0192[.5(-40) + .25(40)]}{1.007} D = -.977\, D$$

(2.a)
$$500\, b_1 + [.5(80) + .25(-40)]\, b_2 = (1)(80) + .5(-40)$$
$$30\, b_1 + \quad 500\, b_2 = .5(80) + .25(-40)$$

$I = .1168\, X_1 + .0530\, X_2 = \hat{G}_{D\alpha}$ and $\sigma_I = 2.932$.

$$(2.b) \qquad \Delta G_D = \frac{b_1 Cov(G_{D\alpha}, X_1) + b_2 Cov(G_{D\alpha}, X_2)}{\sigma_I} D$$

$$= \frac{.1168(60) + .0530(30)}{2.932} D = 2.932\ D$$

$$\Delta G_M = \frac{b_1 Cov(G_{M\alpha}, X_1) + b_2 Cov(G_{M\alpha}, X_2)}{\sigma_I} D$$

$$= \frac{.1168(-20) + .0530(-10)}{2.932} D = -.977\ D$$

JOINT SELECTION FOR THE DIRECT AND MATERNAL GENETIC EFFECTS

For one phenotypic trait assume the overall economic value is determined partly by the direct genetic component and partly by the maternal genetic component so that aggregate genetic economic value for animal α is $T_\alpha = v_D G_{D\alpha} + v_M G_{M\alpha}$ where v_D is the net economic value for the direct contribution and v_M is the net economic value for the maternal contribution. These economic values are not necessarily the same because, although the gross price is the same for the total product, the cost of production may be different for the direct and maternal portions.

The usual selection index procedure for selecting for overall genetic value can be used except that records on at least two relatives are needed because the maternal and direct traits are measured jointly. In addition the two kinds of relatives must be such that $(a_{\alpha 1}, a_\alpha W_1)$ is not proportional to $(a_{\alpha 2}, a_\alpha W_2)$. This restriction will be illustrated later.

The procedure for joint selection for direct and maternal genetic value will be illustrated for one trait and using records on only two relatives, X_1 and X_2. The index will be $I_\alpha = b_1 X_1 + b_2 X_2$ to estimate $T_\alpha = v_D G_{D\alpha} + v_M G_{M\alpha}$.

The general equations which determine the b's are:

$$\sigma^2_{X_1} \quad b_1 + \sigma_{X_1 X_2} b_2 = \sigma_{X_1 T_\alpha}$$
$$\sigma_{X_1 X_2} b_1 + \sigma^2_{X_2} \quad b_2 = \sigma_{X_2 T_\alpha} \ .$$

The coefficients of the b's are the same as for selection for $G_{D\alpha}$ or $G_{M\alpha}$. The covariances between the X's and T_α can be computed as:

$$\sigma_{X_i T_\alpha} = v_D \sigma_{X_i G_{D\alpha}} + v_M \sigma_{X_i G_{M\alpha}}$$
$$= v_D \left(a_{i\alpha} \sigma^2_{G_D} + a_{W_i \alpha} \sigma_{G_D G_M} \right) + v_M \left(a_{i\alpha} \sigma_{G_D G_M} + a_{W_i \alpha} \sigma^2_{G_M} \right).$$

An alternative procedure would be to index $G_{D\alpha}$ and $G_{M\alpha}$ separately and then weight each by v_D and v_M, i.e., $I_\alpha = v_D I_{D\alpha} + v_M I_{M\alpha}$ where $I_{D\alpha} = \hat{G}_{D\alpha}$ and $I_{M\alpha} = \hat{G}_{M\alpha}$. The expected response by selection can be computed as before:

$\Delta T = v_D \, \Delta G_D + v_M \, \Delta G_M$ where

$$\Delta G_D = \frac{COV\left(G_{D\alpha}, I\right)}{\sigma_I} D \quad \text{and} \quad \Delta G_M = \frac{COV\left(G_{M\alpha}, I\right)}{\sigma_I} D \ .$$

The following example problems illustrate the computations for selecting for both the direct and maternal genetic traits. Example 1 illustrates the futility of trying to select for both when $(a_{1\alpha}, a_{W_1\alpha})$ and $(a_{2\alpha}, a_{W_2\alpha})$ are proportional. Example 3 shows the effect of changing the sign of the genetic covariance between the direct and maternal genetic values.

Example Problems Of Selecting For G_D And G_M Simultaneously

Problems

Given: $\sigma^2_{G_D} = 80$, $\sigma^2_{G_M} = 40$, $\sigma_{G_D G_M} = 40$, $\sigma^2_P = 500$

1. Suppose X_1 = record on the sire, X_2 = record on the dam.

   If $v_D = 4$ and $v_M = 1$ can selection be for $T_\alpha = (4)G_{D\alpha} + (1)G_{M\alpha}$?

2. a) Suppose X_1 = record on dam, X_2 = record on a paternal half sib

   What is the index for selecting for $T_\alpha = (4)G_{D\alpha} + (1)G_{M\alpha}$?

   b) What is the expected correlated response in G_D and G_M ?

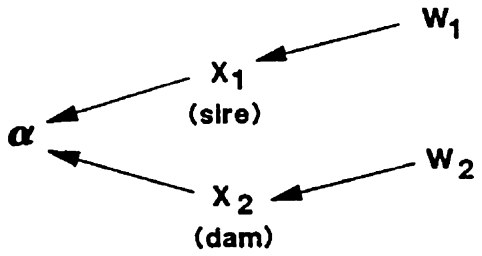
3. Repeat 2a and b when $\sigma_{G_D G_M} = -40$.

Solutions

1. X_1 = record on sire and X_2 = record on dam

   $(a_{1\alpha}, a_{W_1\alpha}) = (.5, .25)$;

   $(a_{2\alpha}, a_{W_2\alpha}) = (.5, .25)$. These are proportional so selection cannot be for $v_D G_D + v_M G_M$.



   For example:

   if $v_D = 4$ and $v_M = 1$, then $I = .46X_1 + .46X_2$ and $\sigma_I = 14.55$ with

$$\Delta G_D = \frac{.46(50) + .46(50)}{14.55} D = 3.16 D$$

$$\Delta G_M = \frac{.46(30) + .46(30)}{14.55} D = 1.90 D$$

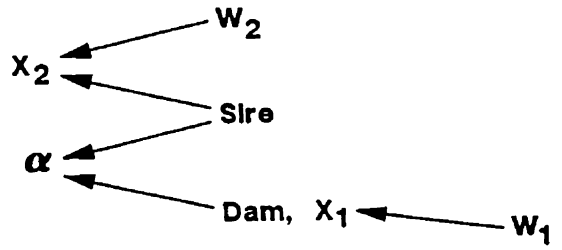
   If $v_D = 1$ and $v_M = 4$, then $I = .34X_1 + .34X_2$ and $\sigma_I = 10.75$ with

$$\Delta G_D = \frac{.34(50) + .34(50)}{10.75} D = 3.16 D$$

$$\Delta G_M = \frac{.34(30) + .34(30)}{10.75} D = 1.90 D$$ as before for $v_D = 4$ and $v_M = 1$.

2. X_1 = record on dam and X_2 = record on a phs

$v_D = 4$ $v_M = 1$



a) $$500b_1 + 0\,b_2 = v_D\left(a_{1\alpha}\sigma^2_{G_D} + a_{W_1\alpha}\sigma_{G_M G_D}\right) + v_M\left(a_{1\alpha}\sigma_{G_M G_D} + a_{W_1\alpha}\sigma^2_{G_M}\right)$$

$$0\,b_1 + 500b_2 = v_D\left(a_{2\alpha}\sigma^2_{G_D} + a_{W_2\alpha}\sigma_{G_M G_D}\right) + v_M\left(a_{2\alpha}\sigma_{G_M G_D} + a_{W_2\alpha}\sigma^2_{G_M}\right)$$

$$500b_1 = 4[.5(80) + .25(40)] + 1[.5(40) + .25(40)] = 230$$

$$500b_2 = 4[.25(80) + 0] + 1[.25(40) + 0] = 90$$

$$I = .46\,X_1 + .18\,X_2 \text{ with } \sigma_I = 11.05.$$

b) $$\Delta G_D = \frac{COV(G_{D\alpha}, I)}{\sigma_I} D = \frac{.46[.5(80) + .25(40)] + .18[.25(80) + 0]}{11.05} D = 2.407\,D$$

$$\Delta G_M = \frac{COV(G_{M\alpha}, I)}{\sigma_I} D = \frac{.46[.5(40) + .25(40)] + .18[.25(40) + 0]}{11.05} D = 1.412\,D$$

3. a) $$500\,b_1 + 0\,b_2 = 4[.5(80) + .25(-40)] + 1[.5(-40) + .25(40)] = 110$$

$$0\,b_1 + 500\,b_2 = 4[.25(80) + 0] + 1[.25(-40) + 0] = 70$$

$$I = .22\,X_1 + .14\,X_2 \text{ with } \sigma_I = 5.83.$$

b) $$\Delta G_D = \frac{.22[.5(80) + .25(-40)] + .14[.25(80) + 0]}{5.83} D = 1.612\,D$$

$$\Delta G_M = \frac{.22[.5(-40) + .25(40)] + .14[.25(-40) + 0]}{5.83} D = -.617\,D.$$

CHAPTER 17

SELECTION WHEN TRAITS INFLUENCED BY GRANDMATERNAL AND MATERNAL EFFECTS

The granddam may, for some traits, affect her daughter's maternal ability which in turn influences the record of the grandprogeny. Beef cattle breeders have reported that cows that were heavy themselves at weaning tend to wean calves that are lighter than cows that were not so heavy at weaning. A grandmaternal effect can be postulated as a cause of this phenomenon.

MODEL WITH MATERNAL AND GRANDMATERNAL EFFECTS

This grandmaternal effect may have a genetic basis in the grandmother (i") but is an environmental effect on the maternal ability of the mother (i') and on the actual phenotype of the calf (i). In fact, the model including maternal effects can be expanded so that the maternal effect is made up of a direct maternal effect and an environmental effect from the grandmother:

$$P_{M_{i'}} = G_{M_{i'}} + E_{i'} = G_{M_{i'}} + G_{N_{i''}} + E_{M_{i'}} + E_{N_{i''}} \quad ,$$

where $G_{M_{i'}}$ is the genetic maternal effect, $G_{N_{i''}}$ is the genetic grandmaternal effect, $E_{M_{i'}}$ is the maternal environmental effect other than that with grandmaternal causes, and $E_{N_{i''}}$ is the nongenetic (environmental) grandmaternal effect.

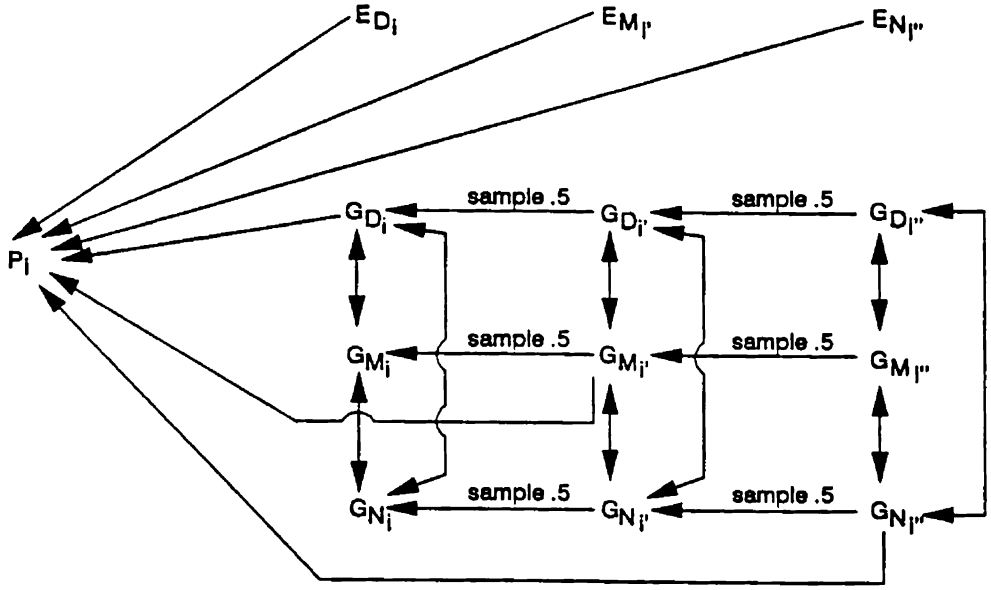
Then, the model for a record on animal i can be partitioned as:

$P_i = G_{D_i} + E_i$ where G_{D_i} is the genetic ability of i,

$P_i = G_{D_i} + P_{M_{i'}} + E_{D_i}$ where $P_{M_{i'}}$ is the total maternal effect of i' on P_i, and

$P_i = G_{D_i} + G_{M_{i'}} + G_{N_{i''}} + E_{D_i} + E_{M_{i'}} + E_{N_{i''}}$.

Only P_i can be measured. The diagram illustrates, as before, that the maternal genetic ability of the mother is expressed only in her progeny. Similarly, the grandmaternal genetic effect is expressed only in the grandprogeny. The double-headed arrows represent a possible covariance due to pleiotropic genetic effects. A sample one-half of the genes for the direct, maternal, and grandmaternal effects are transmitted in each generation from parent to offspring.



COVARIANCES BETWEEN RELATIVES

The covariances among relatives (e.g., X and Y) can be determined as before from $E(P_X P_Y)$ where:

$$P_X = G_{D_X} + G_{M_{X'}} + G_{N_{X''}} + E_{D_X} + E_{M_{X'}} + E_{N_{X''}} \text{ and}$$

$$P_Y = G_{D_Y} + G_{M_{Y'}} + G_{N_{Y''}} + E_{D_Y} + E_{M_{Y'}} + E_{N_{Y''}} .$$

To simplify the expectation, consider only $G_X = G_{D_X} + G_{M_{X'}} + G_{N_{X''}}$ and $G_Y = G_{D_Y} + G_{M_{Y'}} + G_{N_{Y''}}$ where G_{D_X} is the genetic value of X for the measured trait, $G_{M_{X'}}$ is the maternal genetic value of the dam of X, X', and $G_{N_{X''}}$ is the grandmaternal genetic value of the maternal granddam of X, X". Thus;

$$E(G_XG_Y) = Cov(G_X,G_Y) = Cov(G_{D_X},G_{D_Y}) + Cov(G_{D_X},G_{M_{Y'}}) + Cov(G_{D_X},G_{N_{Y''}}) +$$
$$Cov(G_{M_{X''}},G_{D_Y}) + Cov(G_{M_{X''}},G_{M_{Y'}}) + Cov(G_{M_{X''}},G_{N_{Y''}}) +$$
$$Cov(G_{N_{X'''}},G_{D_Y}) + Cov(G_{N_{X'''}},G_{M_{Y'}}) + Cov(G_{N_{X'''}},G_{N_{Y''}}).$$

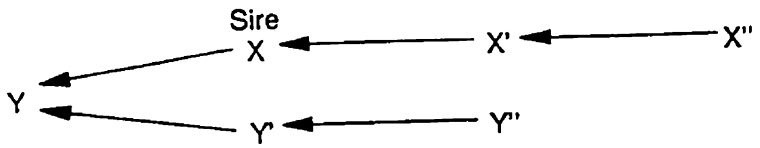
Each of these terms can be evaluated in terms of additive, dominance, additive by additive, etc., components of genetic variance and covariance (where the direct, maternal, and grandmaternal components are considered separate traits).

If only additive genetic effects are assumed, then:

$$Cov(G_X,G_Y) = a_{XY}\sigma^2_{G_D} + (a_{XY'} + a_{YX'})\sigma_{G_D}G_M + (a_{XY''} + a_{YX''})\sigma_{G_D}G_N +$$
$$a_{X'Y'}\sigma^2_{G_M} + (a_{X'Y''} + a_{Y'X''})\sigma_{G_M}G_N + a_{X''Y''}\sigma^2_{G_N},$$

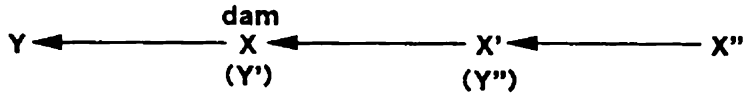
where the variances are additive genetic variances and the covariances are additive genetic covariances among the direct (D), maternal (M), and grandmaternal (N) effects.

The necessary additive relationships can be found from careful drawing of the pedigree of symbolic animals, X and Y; their dams, X' and Y'; and maternal granddams, X" and Y". For example, if X is a sire and Y is the progeny, the diagram is:



Thus, $a_{XY} = .5$, $a_{XY'} = 0$, $a_{YX'} = .25$, $a_{XY''} = 0$, $a_{YX''} = .125$, $a_{X'Y'} = 0$, $a_{X'Y''} = 0$, $a_{Y'X''} = 0$, and $a_{X''Y''} = 0$.

However, if X is a dam and Y is the progeny, the relationships are different:



Now, $a_{XY} = .5$, $a_{XY'} = 1$, $a_{YX'} = .25$, $a_{XY"} = .5$, $a_{YX"} = .125$, $a_{X'Y'} = .5$, $a_{X'Y"} = 1$, $a_{Y'X"} = .25$, and $a_{X"Y"} = .5$. The relationships which are coefficients of the variances and covariances for some common relatives are given below.

	Component					
Relatives	$\sigma^2_{G_D}$	$\sigma_{G_D G_M}$	$\sigma_{G_D G_N}$	$\sigma^2_{G_M}$	$\sigma_{G_M G_N}$	$\sigma^2_{G_N}$
	Coefficient					
P_X, P_Y	a_{XY}	$a_{XY'} + a_{YX'}$	$a_{XY"} + a_{YX"}$	$a_{X'Y'}$	$a_{X'Y"} + a_{Y'X"}$	$a_{X"Y"}$
With self	1	1	.500	1	1	1
Sire, progeny	.500	.250	.125	0	0	0
Dam, progeny	.500	1.250	.625	.500	1.250	.500
Full sibs	.500	1	.500	1	1	1
Maternal sibs	.250	1	.500	1	1	1
Paternal sibs	.250	0	0	0	0	0
Granddam, grandprogeny	.250	.625	1.062	.250	.625	.250

SELECTION INDEX EQUATIONS

The problem of selection is similar to that in the presence of maternal effects. The selection index equations are modified to take into account the direct, maternal, and grandmaternal components.

If, for example, selection is for $T = G_{D\alpha}$ (additive direct for animal α) then the right-hand sides become:

$$\sigma_{X_i T} = E[(G_{D X_i} + G_{M X'_i} + G_{N X''_i} + \text{other } E_{X_i})(G_{D\alpha})]$$
$$= a_{i\alpha}\sigma^2_{G_D} + a_{i'\alpha}\sigma_{G_D G_M} + a_{i''\alpha}\sigma_{G_D G_N} \text{ and } \sigma^2_T = \sigma^2_{G_D} .$$

If $T = G_{M\alpha}$, then:

$$\sigma_{X_i T} = a_{i\alpha}\sigma_{G_D G_M} + a_{i'\alpha}\sigma^2_{G_M} + a_{i''\alpha}\sigma_{G_M G_N} \text{ and } \sigma^2_T = \sigma^2_{G_M} .$$

If $T = G_{N\alpha}$, then:

$$\sigma_{X_i T} = a_{i\alpha}\sigma_{G_D G_N} + a_{i'\alpha}\sigma_{G_M G_N} + a_{i''\alpha}\sigma^2_{G_N} \text{ and } \sigma^2_T = \sigma^2_{G_N} .$$

If some function $T = v_D G_{D\alpha} + v_M G_{M\alpha} + v_N G_{N\alpha}$ is the overall merit where the v's are economic values of the components, then:

$$\hat{T} = v_D \hat{G}_{D\alpha} + v_M \hat{G}_{M\alpha} + v_N \hat{G}_{N\alpha} .$$

The selection index weights also can be determined directly using the RHS's:

$$\sigma_{X_i T} = E[(G_{D X_i} + G_{M X'_i} + G_{N X''_i} + \text{other E's})(v_D G_{D\alpha} + v_M G_{M\alpha} + v_N G_{N\alpha})]$$
$$= v_D(a_{i\alpha}\sigma^2_{G_D} + a_{i'\alpha}\sigma_{G_D G_M} + a_{i''\alpha}\sigma_{G_D G_N}) +$$
$$v_M(a_{i\alpha}\sigma_{G_D G_M} + a_{i'\alpha}\sigma^2_{G_M} + a_{i''\alpha}\sigma_{G_M G_N}) +$$
$$v_N(a_{i\alpha}\sigma_{G_D G_N} + a_{i'\alpha}\sigma_{G_M G_N} + a_{i''\alpha}\sigma^2_{G_N})$$

and $\sigma^2_T = E[(v_D G_{D\alpha} + v_M G_{M\alpha} + v_N G_{N\alpha})^2]$.

Records on at least three kinds of relatives (where the three additive relationships $a_{i\alpha}$, $a_{i'\alpha}$ and $a_{i''\alpha}$ are not proportional for the three relatives) are necessary for selection with different economic values for the direct, maternal, and grandmaternal components.

CHAPTER 18

FETAL EFFECTS MODEL (SIRE OF FETUS EFFECT)

Some traits of a female may be influenced by the fetus she is carrying either during gestation or following gestation. An obvious example is the ease with which the mother gives birth. The genes of the mother directly affect ease of birth, but the size of the fetus also may affect the ease of birth by its mother. The size of the fetus is certainly partially influenced by the genes it carries. In some species there is speculation that hormones secreted by the fetus may influence the development of secretory tissue and thus influence milk production during the last part of gestation or during lactation which follows birth of the fetus.

The fetal effects model is similar to the maternal effects model except that the embedded trait is a property of the fetus the animal is carrying rather than of the mother of the animal. Figure 18.1 shows that the animal making the record contributes a sample one-half of the fetal genes as does the sire of the fetus. If these genes contribute to the fetal effect (the embedded trait) then the sire, through those genetic effects, can influence the performance of his unrelated mate. The effect has been called the sire of fetus effect or the service sire effect.

FETAL EFFECTS MODEL

Figure 18.1 shows the genetic and environmental components for both the direct and fetal effects on records of relatives x and y. The fetal effect could be on the current record or on a subsequent record. The same model applies to calving difficulty. In fact, any trait that is influenced by the mate of the female can be described by such a model. Fixed effects on the records will be ignored here but would need to be considered in prediction procedures or in estimation of components of variance.

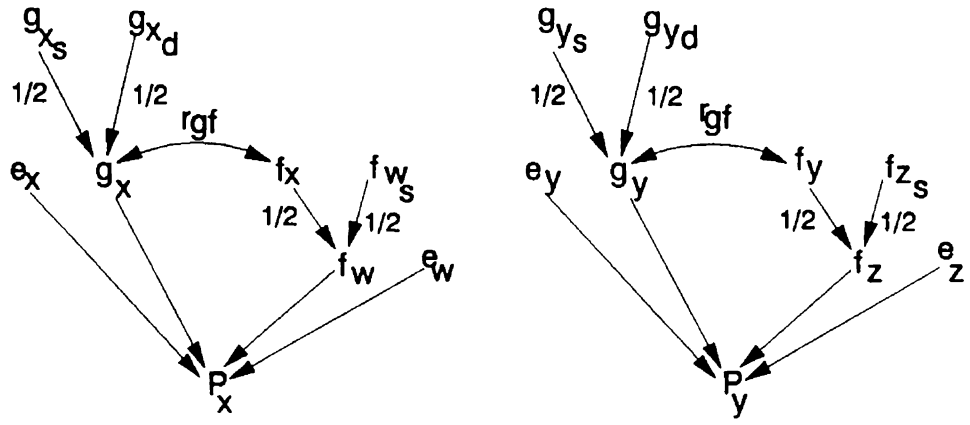


Figure 18.1. Diagram of direct genetic and environmental effects (g_x and e_x) and fetal genetic and environmental effects (f_w and e_w) on the phenotypic record of animal, x, carrying fetus, w. The sire and dam of x are x_s and x_d, w_s is the sire of the fetus, and x is the dam of the fetus. A similar diagram is given for any potential relative, y, carrying fetus, z.

Linear models including effects shown in Figure 18.1 were presented by Willham (1963). His application was to a maternal effects model.

The fetal effects model is:

$$P_x = g_x + f_w + e_x + e_w \text{ and } P_y = g_y + f_z + e_y + e_z$$

where the **g**'s are genetic values for the direct effect on P, the **f**'s are the genetic effects of the fetus on P, and the **e**'s are corresponding environmental effects. The pair of animals with records are x and y; w and z are the fetuses having sires, w_s and z_s. In the usual P = G + E model, all the effects except g would be included in E. The **f** effects are environmental to the animal making the record but are, in part, genetically determined. The covariance is: $Cov(P_x,P_y) = Cov(g_x,g_y) + Cov(f_w,f_z) + Cov(g_x,f_z) + Cov(g_y,f_w)$. If only additive genetic effects are considered or are assumed important, the covariances can be written as by Willham (1963) :

$$Cov(P_x,P_y) = a_{xy}\sigma_g^2 + a_{wz}\sigma_f^2 + (a_{xz} + a_{yw})\sigma_{gf}$$

where the a's are additive or numerator relationships. If **f** is a fetal effect, then σ_g^2 is the variance of direct additive genetic effects, σ_f^2 is the variance of additive fetal genetic effects, and σ_{gf} is the covariance between additive direct and additive fetal genetic effects.

COVARIANCES BETWEEN RELATIVES

The previous expression can be used to determine the theoretical covariance between records of any pair of relatives, x and y, when influenced by fetuses of sires, w_s and z_s. For example, when calculating the covariance between records of a dam and her daughter when the dam's record was made with the influence of the fetus (her daughter), x is the daughter, $y = x_d$ is the dam, x_s is the sire of x but is also z_s, the sire of the fetus, x.

Then:

$$\text{Cov(daughter,dam)} = .5\,\sigma_g^2 + .5\,\sigma_f^2 + (1 + .25)\sigma_{gf}\,.$$

Expectations of covariances between typical pairs of records are given in Table 18.1.

TABLE 18.1. COEFFICIENTS OF THE ADDITIVE GENETIC VARIANCES FOR THE DIRECT EFFECT, σ_g^2, FOR THE FETAL EFFECT, σ_f^2, AND OF THE ADDITIVE GENETIC COVARIANCE BETWEEN THE DIRECT AND FETAL EFFECTS, σ_{gf}, FOR THE COVARIANCE[a] BETWEEN VARIOUS RELATIVES AND COMBINATIONS OF SIRES OF FETUSES

Animals with records	Sire of fetus	Coefficient of σ_g^2	Coefficient of σ_f^2	Coefficient of σ_{gf}
Daughter, dam	Daughter not from service sire of dam	.500	.1250	.500
Daughter, dam	Daughter from service sire of dam	.500	.5000	1.250
Full sibs	Different	.500	.1250	.500
Full sibs	Same	.500	.3750	.500
Paternal or maternal sibs	Different	.250	.0625	.250
Paternal or maternal sibs	Same	.250	.3125	.250
Maternal sibs	Sire of x is service sire of y	.250	.1875	.500
Unrelated	Same	0	.2500	0

[a]Note that these covariances may also include other components due to effects such as direct, dominance and maternal, additive genetic effects.

The practical implications are that the effect of the sire of a daughter includes the value of the sample one-half of his genes concerned directly with her production and a sample one-quarter of his genes associated with the fetal effect since he is the grandsire of every calf his daughter produces. Thus, the sire of daughter effect (which is thought of as the sire transmitting ability in absence of fetal effects) is:

g/2 for production of daughter + f/4 for production of daughter.

The mate effect, or the fetal effect of the sample one-half of the genes contributed by the mating sire to the fetus, is expressed in the lactation performance of the mother:

Sire of fetus effect (mate of dam) = f/2 for production of dam (his mate).

Note that the sire of fetus is also the sire of the replacement female, resulting from birth and survival of the fetus. Thus, in the next generation, the sire of the fetus becomes the sire of the daughter. If there is a negative relationship between direct and fetal effects, then effective selection may be difficult. The other dilemma is that even if the effects are unrelated, should more emphasis be placed on selection of a sire for his fetal effect, which almost immediately influences production of the mate, or on selection for his direct genetic value, which does not become expressed until the resulting offspring become productive?

SELECTION INDEX EQUATIONS

Selection index weights can be found as usual for selection for direct or fetal effects by modifying the coefficients on the LHS according to covariances such as those in Table 18.1. The RHS can be found by the rules for expected values using the model for the records that includes fetal effects. These principles were demonstrated in Chapters 16 and 17 for models including maternal and grandpaternal effects.

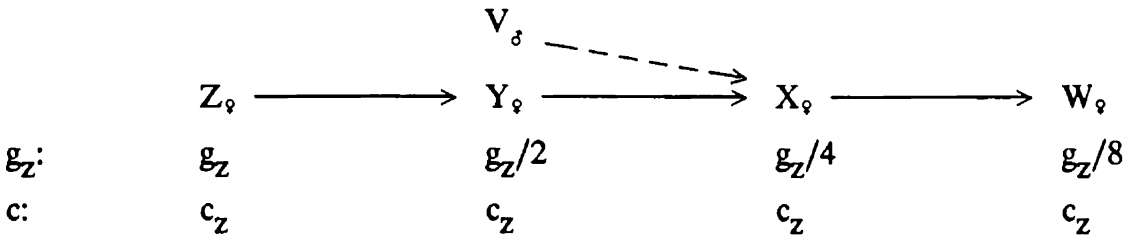
CHAPTER 19

CYTOPLASMIC EFFECTS MODEL

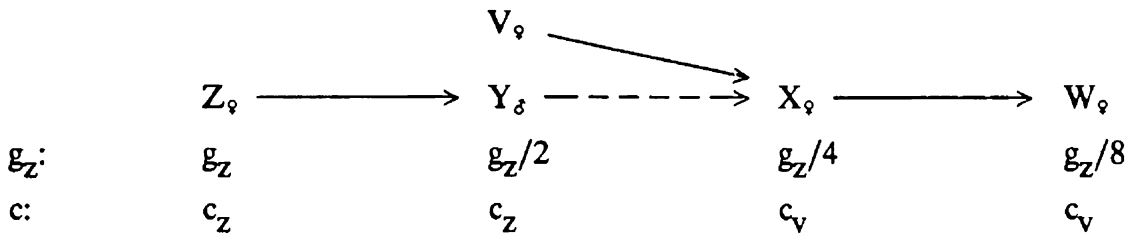
Cytoplasm of the fertilized ovum comes primarily from the mother. Mitochondria in the cytoplasm are responsible for cellular metabolism. The DNA of mitochondria in most species is inherited primarily or entirely from the mother. Thus, cytoplasmic effects generally are considered to be maternal in origin and essentially to be unchanging along the maternal line (Figure 19.1). Males express the cytoplasmic effects received from their mothers but will not transmit their cytoplasm to their offspring (Figure 19.1).

FIGURE 19.1 FRACTION OF ADDITIVE GENETIC (g) AND CYTOPLASMIC (c) EFFECTS IN DESCENDANTS.

For female line of descent:



With male in line of descent:



Cytoplasmic effects can be incorporated easily into selection index procedures either in computing the variances and covariances among the X's or in computing the right-hand sides for selection of a function of direct, additive genetic and cytoplasmic effects.

For the purpose of illustration, assume the only genetic effects other than cytoplasmic effects are additive direct effects. Maternal effects which may be confounded with cytoplasmic effects can be included in the model easily.

COVARIANCES FOR MODELS WITH CYTOPLASMIC EFFECTS

Models with cytoplasmic effects for records after adjustment for fixed factors on relatives x and y are:

$$P_x = g_x + c_f + b_{x_f} + e_x \text{ and } P_y = g_y + c_{f'} + b_{y_{f'}} + e_y$$

where **g** is the additive genetic value for the direct effect on phenotype, **c** is the cytoplasmic effect originating in the female line with animal **f** or **f'**, **b** is the interaction between additive genetic and cytoplasmic effects, and the **e**'s are random and independent environmental effects (may include random cytoplasmic effects). Then when covariances between **g**'s and **c**'s, **g**'s and **b**'s, **g**'s and **e**'s, **c**'s and **b**'s, **c**'s and **e**'s, and **b**'s and **e**'s are zero:

$$\text{Cov}(P_x,P_y) = \text{Cov}(g_x,g_y) + \text{Cov}(c_f,c_{f'}) + \text{Cov}(b_{x_f}, b_{y_{f'}}) + \text{Cov}(e_x,e_y) \ .$$

If f = f': $\text{Cov}(c_f,c_{f'}) = \sigma_c^2$ and = 0 otherwise, and

$\text{Cov}(b_{x_f}, b_{y_{f'}}) = a_{xy}\sigma_b^2$ and 0 otherwise .

Let $c_{xy}\sigma_x^2$ be the environmental covariance between records of **x** and **y**.

Thus for f = f' : $\text{Cov}(P_x, P_y) = a_{xy}\sigma_g^2 + \sigma_c^2 + a_{xy}\sigma_b^2 + c_{xy}\sigma_x^2$.

And for f ≠ f' : $\text{Cov}(P_x, P_y) = a_{xy}\sigma_g^2 + c_{xy}\sigma_x^2$.

The table gives the expected composition of covariances of common relatives.

Relationship	Component σ_g^2	Component σ_c^2	Component σ_b^2	Environmental covariance $/\sigma_x^2$
Female parent, offspring	.50	1	.50	$c_{FP,O}$
Male parent, offspring	.50	0	0	$c_{MP,O}$
Maternal half sibs	.25	1	.25	c_{MHS}
Paternal half sibs	.25	0	0	c_{PHS}
Full sibs	.50	1	.50	c_{FS}
Female grandparent, offspring	.25	1	.25	$c_{FG,O}$
Animal with self	1	1	1	1
Identical twins	1	1	1	c_{IT}
Unrelated nuclei in same cytoplasm	0	1	0	c_{NC}

If σ_g^2, σ_c^2, σ_b^2, and c_{xy} are known, then variances and covariances can be calculated for the coefficients of the selection index equations to find the selection index weights.

RIGHT HAND SIDES FOR SELECTION INDEX EQUATIONS

If selection is for direct additive genetic value of animal **α**, then the right-hand sides of the selection index equations as usual will be:

$$\sigma_{X_i T} = a_{i\alpha}\sigma_g^2$$

where $a_{i\alpha}$ is the additive relationship between **α** and **i**.

If selection is for direct additive genetic value of animal α plus the cytoplasmic value of α plus the interaction, $T = g_\alpha + c_\alpha + b_{\alpha f'}$, then:

$$\text{if } f = f' \;,\quad \sigma_{X_i T} = a_{i\alpha}\sigma_g^2 + \sigma_c^2 + a_{i\alpha}\sigma_b^2 \quad \text{and}$$

$$\text{if } f \neq f' \;,\quad \sigma_{X_i T} = a_{i\alpha}\sigma_g^2 \;.$$

In general, for $T = g_\alpha + c_\alpha + b_{\alpha f'}$; $\sigma_{X_i T} = a_{i\alpha}\sigma_g^2 + P(f=f')[\sigma_c^2 + a_{i\alpha}\sigma_b^2]$ where $P(f=f')$ is the probability that the cytoplasm of relative **i** with the record X_i and the cytoplasm of the animal being evaluated, α, is the same.

If $T = g_\alpha + c_\alpha$, and then:

$$\text{if } f = f' \;,\; \sigma_{x_i T} = a_{i\alpha}\sigma_g^2 + \sigma_c^2 \quad \text{and}$$

$$\text{if } f \neq f' \;;\; \sigma_{x_i T} = a_{i\alpha}\sigma_g^2 \;.$$

Unless σ_c^2 is relatively large, selection for direct additive genetic value while ignoring cytoplasmic effects is likely to be nearly as effective as jointly selecting for direct additive and cytoplasmic effects.

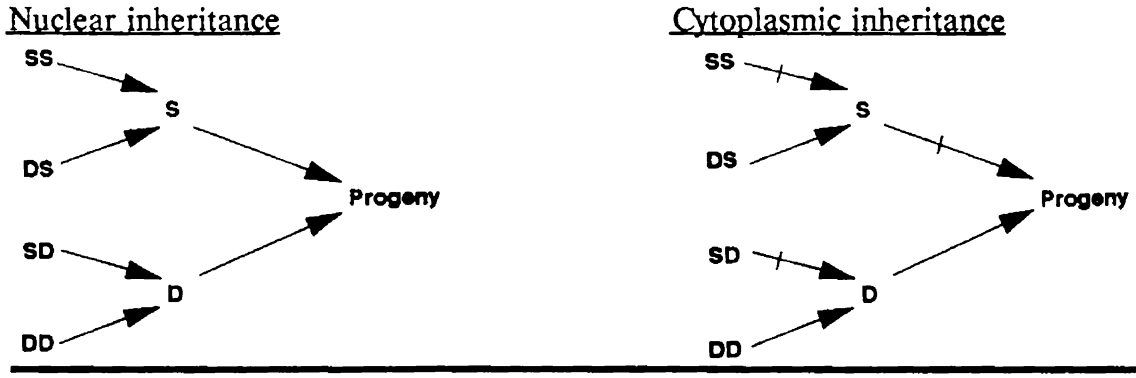
BIAS IN HERITABILITY ESTIMATES

Heritability (additive direct) can be overestimated from covariances between relatives with the same cytoplasm if cytoplasmic effects on the trait are real and if those effects are ignored. Overestimates of heritability will lead to overestimates of the accuracy, r_{TI}, of evaluation and overestimation of expected superiority for additive genetic value from selection because both r_{TI} and σ_T will be overestimated.

EXPECTED RESPONSE TO SELECTION

Selection for cytoplasmic genetic value in addition to direct additive genetic value can be relatively important to total genetic gain only when the reproductive rate of females to produce female replacements is greatly increased. The reason is that of the four paths of selection (in the case of milk yield in dairy cows), cytoplasmic effects are transmitted only through the dam to female path.

FIGURE 19. 2. THE FOUR PATHS OF SELECTION



For sire to sire, dam to sire, and sire to dam paths, selection should be for additive genetic value with selection differentials of ΔSS, ΔDS, and ΔSD.

For dam to dam path, selection can be for sum of direct additive and cytoplasmic effects with the selection differential partitioned into ΔDD_g (direct additive) and ΔDD_c (cytoplasmic). These two parts can be obtained theoretically by calculation of expected correlated response. If I is the index for the sum, g + c, then:

$$\Delta DD_g = \frac{[Cov(g,I)]}{\sigma_I} D \quad \text{and} \quad \Delta DD_c = \frac{[Cov(c,I)]}{\sigma_I} D$$

where D is the standardized selection intensity factor. Note that D and σ_I are the same for both calculations. With no covariance between g and c, the only contributions to Cov(c,I) will be from females in direct female line of descent such as daughter, dam, maternal granddam, maternal half sisters, and full sibs.

The direct additive genetic differential applies to the usual formula for genetic gain from four paths of selection.

The differential due to cytoplasmic effects contributes immediately to progeny and thus gain per year for cytoplasmic effects is the cytoplasmic differential divided by the generation interval for the dam of dam path. (Some scientists have reasoned that since females to be dams of dams are selected jointly for direct and cytoplasmic effects, the division should be by the sum of generation intervals.) Thus on a per year basis:

$$\Delta(g+c) = \frac{\Delta SS + \Delta DS + \Delta SD + \Delta DD_g}{L_{SS} + L_{DS} + L_{SD} + L_{DD}} + \frac{\Delta DD_c}{L_{DD}} .$$

Because increased reproductive rate in females results in the same increase in selection intensity for ΔDD_g and ΔDD_c, the equation can be partitioned into the three paths, Δg_3, the sum of paths that do not contribute cytoplasm to the population and the two parts due to dams of dams:

$$\Delta(g+c) = \Delta g_3 + \frac{\Delta DD_g}{\Sigma L} + \frac{\Delta DD_c}{L_{DD}} .$$

For example, if σ_c^2 is 5% and σ_g^2 is 25% of the phenotypic variance for production of dairy cattle, the gain per year from increasing the standardized selection intensity factor will be somewhat greater from $\Delta DD_g/\Sigma L$ than from $\Delta DD_c/L_{DD}$ even though L_{DD} is only about one-fourth of ΣL. The extra gain due to ΔDD_c can be substantial if σ_c^2 is as great as 5% of phenotypic variance and replacement females can be obtained from the top 10 to 50% of the herd. Such an increase in reproductive rate would require sexing of semen or multiple ovulation and embryo transfer. The costs of those reproductive systems must be balanced against the value of the additional genetic gain.

Whether cytoplasmic effects can account for as much as 5% of variation is doubtful. Because cytoplasmic effects seem to be transmitted essentially as a whole, segregation and recombination are not available to maintain variability. Thus, cytoplasmic lines may soon be fixed because selection should be relatively effective. Not many combinations of mitochondrial DNA would be expected as compared to the combinations of nuclear DNA. The few combinations of mitochondrial DNA that do survive after a number of generations of selection may all be nearly optimum for effects on production or reproduction.

APPENDIX TO CHAPTER 19

COVARIANCE BETWEEN RELATIVES WITH SINGLE LOCUS FOR ADDITIVE EFFECTS AND CYTOPLASMIC EFFECTS

Let records of relatives **x** and **y** be represented as:

$$x_{ijt} = \alpha_i + \alpha_j + \tau_t + (\alpha\tau)_{it} + (\alpha\tau)_{jt} + e_x$$

$$y_{k\ell u} = \alpha_k + \alpha_\ell + \tau_u + (\alpha\tau)_{ku} + (\alpha\tau)_{\ell u} + e_y$$

where each α_m represents an additive genetic effect of gene **m**,

τ_n represents a cytoplasmic effect of cytoplasm **n**,

$(\alpha\tau)_{mn}$ represents the interaction of the m^{th} additive effect and n^{th} cytoplasmic effect, and

e_w represents environmental effects.

Note: additive genetic value, $g_{ij} = \alpha_i + \alpha_j$ with $\sigma_g^2 = E[\alpha_i^2] + E[\alpha_j^2]$;

cytoplasmic value, $c_t = \tau_t$ with $\sigma_c^2 = E[\tau_t^2]$;

genetic by cytoplasmic interaction, $b_{ijt} = (\alpha\tau)_{it} + (\alpha\tau)_{jt}$

with $\sigma_b^2 = E[(\alpha\tau)_{it}^2] + E[(\alpha\tau)_{jt}^2]$ and

by assumption; $E[g_{ij}c_t] = 0$, $E[g_{ij}b_{ijt}] = 0$, $E[c_t b_{ijt}] = 0$.

Note: $P(i=k) + P(i=\ell) + P(j=k) + P(j=\ell) = 2a_{xy}$ and $P(i=k) = a_{xy}/2$.

Let $P(t=u)$ be the probability that the cytoplasm of **x** is the same as the cytoplasm of **y** (probability is either 1 or 0).

$Cov(x_{ijt}, y_{k\ell u})$:

$$E[g_{ij}g_{k\ell}] = E[(\alpha_i + \alpha_j)(\alpha_k + \alpha_\ell)] = E[\alpha_i\alpha_k + \alpha_i\alpha_\ell + \alpha_j\alpha_k + \alpha_j\alpha_\ell].$$

But $E[\alpha_i\alpha_k] = (a_{xy}/2)E[\alpha^2] = (a_{xy}/4)\sigma_g^2$.

Thus, $E[g_{ij}g_{k\ell}] = a_{xy}\sigma_g^2$.

$E[c_t c_u) = E[\tau_t \tau_u] = P(t=u)\sigma_c^2$; that is, either σ_c^2 or 0.

$$E[b_{ijt}b_{k\ell u}] = E\{[(\alpha\tau)_{it} + (\alpha\tau)_{jt}][(\alpha\tau)_{ku} + (\alpha\tau)_{\ell u}]\}$$

$$= E[(\alpha\tau)_{it}(\alpha\tau)_{ku} + (\alpha\tau)_{it}(\alpha\tau)_{\ell u} + (\alpha\tau)_{jt}(\alpha\tau)_{ku} + (\alpha\tau)_{jt}(\alpha\tau)_{\ell u}].$$

But for $t = u$; $E[(\alpha\tau)_i(\alpha\tau)_k] = (a_{xy}/2)\,E[(\alpha\tau)^2] = (a_{xy}/4)\sigma_b^2$.

Thus, $E[b_{ijt}b_{k\ell u}] = P(t=u)a_{xy}\sigma_b^2$; that is, either $a_{xy}\sigma_b^2$ or 0.

Therefore,

$$Cov(x,y) = a_{xy}\sigma_g^2 + P(t=u)\sigma_c^2 + P(t=u)a_{xy}\sigma_b^2.$$

CHAPTER 20

SELECTION FOR TRAITS WITH NONLINEAR ECONOMIC VALUE

Two general problems not covered by the usual selection index procedure involve: (1) the situation where the value of the product changes with the output of the product, e.g., the value of an additional pound of milk when the level is 109 lb per day is not the same as when the level is 19 lb per day, and (2) the situation where the value of a trait depends on the level of another trait, e.g., the value of milk depends on the fat test of the milk.

NONLINEAR MERIT

If costs and income for production are known for different levels of production for some trait, the net income curve may be approximated by some nonlinear or polynomial function, e.g.,

Net income = $c + v_1(X_1+\mu_1) + v_2(X_1+\mu_1)^2 + v_3(X_1+\mu_1)^3 + \cdots$, where c is a constant, the v's are the appropriate polynomial regression coefficients from fitting net income to polynomials in total yield, $X_1 + \mu_1$, where μ_1 is a population constant and X_1 is the phenotypic deviation from μ_1.

Thus, net genetic merit might be defined as:

$T = c + v_1(G_1+\mu_1) + v_2(G_1+\mu_1)^2 + v_3(G_1+\mu_1)^3 + \cdots$, where G_1 is the

usual additive genetic value for trait 1. The net genetic merit will depend on μ_1 as well as G_1. Animals could rank differently in populations with different average levels of production, μ_1.

A possible procedure for use in selection for net genetic value is to estimate G_1 as usual by I_1 and substitute it into the economic equation so that

$$\hat{T} = I = c + v_1(I_1+\mu_1) + v_2(I_1+\mu_1)^2 + v_3(I_1+\mu_1)^3 + \cdots .$$

If only v_1 and v_2 are nonzero (linear and quadratic values), then this is an optimum procedure for minimizing $E(T - \hat{T})^2$ except for a constant. This has been called the quadratic index. The procedure may be nearly optimum for other cases although for the cubic case Mao and Henderson (personal communications) have shown mathematically that substituting I_1 for G_1 is not identical to finding an index by minimizing $E(T - \hat{T})^2$.

This concept can be extended to more than one trait and to cases where levels of one trait determine the value of another trait. As long as terms in the economic equation are no higher degree than $(X_1+\mu_1)^2$ or $(X_1+\mu_1)(X_2+\mu_2)$, the procedure of substituting the index for each trait into the economic equation is optimum.

For example, with two traits, if

$$T = c + v_1(\mu_1+G_1) + v_2(\mu_2+G_2) + v_3(\mu_1+G_1)(\mu_2+G_2) + v_4(\mu_1+G_1)^2 + v_5(\mu_2+G_2)^2,$$

then the best index for T where $I_1 = \hat{G}_1$ and $I_2 = \hat{G}_2$ is:

$$\hat{T} = c' + v_1(\mu_1+I_1) + v_2(\mu_2+I_2) + v_3(\mu_1+I_1)(\mu_2+I_2) + v_4(\mu_1+I_1)^2 + v_5(\mu_2+I_2)^2,$$

where c' is a constant for all $\hat{T}$. Wilton showed that this is equivalent to

$$I = c + ß_1X_1 + ß_2X_2 + ß_3X_1X_2 + ß_4X_1^2 + ß_5X_2^2$$

where the ß's are the solutions to the equations obtained from minimizing:

$$E[(T\text{-}I) - E(T\text{-}I)]^2.$$

A special example is in the pricing of milk where the value of milk depends on the level of fat test. The example does not consider any other nonlinear economic value for milk. The income equation for milk can be written as:

Income = $(\mu_1+X_1)[v_m + v_f (\mu_2+X_2\text{-base test})]$, where $\mu_1 + X_1$ is the milk record, $\mu_2 + X_2$ is the fat test, v_m is the base price of milk per lb when the milk has the base test, and v_f is the differential in price of milk for a change in fat test. The equation can be rewritten to compare with the quadratic income equation as:

Income = $[v_m - v_f (\text{base test})](\mu_1+X_1) + v_f (\mu_1+X_1)(\mu_2+X_2)$.

Thus, the best index is:

$I = [v_m - v_f (\text{base test})](\mu_1+I_1) + v_f (\mu_1+I_1)(\mu_2+I_2)$,

where $v_1 = v_m - v_f$ (base test) and $v_3 = v_f$ and

$I_1 = b_{11}X_1 + b_{12}X_2$ and $I_2 = b_{21}X_1 + b_{22}X_2$.

In some cases a simpler approximation of I_1 or I_2 may be substituted especially when I_1 and I_2 are based on many progeny, i.e., I_1 may include only records for trait 1 on many progeny. In all cases with a quadratic index, correlated responses are difficult to compute because of terms such as $E(X_1^2X_2)$.

EXAMPLE OF SELECTION WHEN MILK PRICE DEPENDS ON FAT TEST

The example demonstrates that an animal that ranks higher in one herd may not in another herd depending on the average milk yield and fat test in the two herds.

Two sires have been evaluated for milk, I_1, and test, I_2. The two sires are to be used in two herds with widely different average milk and test.

Sire	I_1	I_2	Herd	μ_1	μ_2
A	+2000 lb	-.003	1	12,000 lb	.040
B	+1000 lb	+.003	2	18,000 lb	.035

Three pricing systems are compared where v_m is the base price per lb of milk at a base test of .035 and v_f is the fat differential--the change in price per lb of milk if fat content changes from none to all. The table gives the economic indexes for the six combinations.

	v_m=.05, v_f=.6		v_m=.05, v_f=.8		v_m=.06, v_f=.4	
Sire	Herd 1	Herd 2	Herd 1	Herd 2	Herd 1	Herd 2
A	$717	$964	$722	$952	$951	$1176
B	712	984	733	996	922	1163

GENERAL PROCEDURE FOR PREDICTING QUADRATIC MERIT

As an example with only two traits assume that overall quadratic merit can be defined as:

$$T = v_1(\mu_1+T_1) + v_2(\mu_2+T_2) + v_{12}(\mu_1+T_1)(\mu_2+T_2) + v_{11}(\mu_1+T_1)^2 + v_{22}(\mu_2+T_2)^2,$$

where the v's are economic values for linear, product, and squared increases in true value for traits 1 and 2 having means μ_1 and μ_2. T_1 and T_2 will have zero means and variances $\sigma^2_{T_1}$ and $\sigma^2_{T_2}$.

T can be rewritten as

$$T = a_0 + a_1T_1 + a_2T_2 + a_{12}T_1T_2 + a_{11}T_1^2 + a_{22}T_2^2,$$

where the constants are: $a_0 = v_1\mu_1 + v_2\mu_2 + v_{12}\mu_1\mu_2 + v_{11}\mu_1^2 + v_{22}\mu_2^2$, $a_1 = v_1 + v_{12}\mu_2 + 2v_{11}\mu_1$, $a_2 = v_2 + v_{12}\mu_1 + 2v_{22}\mu_2$, $a_{11} = v_{11}$, $a_{12} = v_{12}$ and $a_{22} = v_{22}$.

Henderson has shown that the best unbiased predictor of T is $\hat{T} = c + I$ where the indexes for traits 1 and 2, I_1 and I_2, are substituted into the quadratic merit equation, $I = a_0 + a_1I_1 + a_2I_2 + a_{12}I_1I_2 + a_{11}I_1^2 + a_{22}I_2^2$, and $c = E(T) - E(I)$. The constant, c, is the same for all animals and therefore will not change ranking and is necessary only to have unbiased predictions.

Only one type of term in $c = E(T) - E(I)$ is difficult to evaluate:

$$\begin{aligned} E(T) &= E(a_0 + a_1T_1 + a_2T_2 + a_{12}T_1T_2 + a_{11}T_1^2 + a_{22}T_2^2) \\ &= a_0 + 0 + 0 + a_{12}\sigma_{T_1T_2} + a_{11}\sigma^2_{T_1} + a_{22}\sigma^2_{T_2}\ ; \\ E(I) &= E(a_0 + a_1I_1 + a_2I_2 + a_{12}I_1I_2 + a_{11}I_1^2 + a_{22}I_2^2) \\ &= a_0 + 0 + 0 + a_{12}E(I_1I_2) + a_{11}\sigma^2_{I_1} + a_{22}\sigma^2_{I_2}\ . \end{aligned}$$

$\sigma^2_{I_1} = r^2_{T_1I_1}\sigma^2_{T_1}$ and $\sigma^2_{I_2} = r^2_{T_2I_2}\sigma^2_{T_2}$ as before where $r^2_{T_iI_i}$ is the squared correlation between T_i and the index prediction, I_i. Thus,

$$E(T) - E(I) = a_{12}[\sigma_{T_1T_2} - E(I_1I_2)] + a_{11}\sigma^2_{T_1}(1-r^2_{T_1I_1}) + a_{22}\sigma^2_{T_2}(1-r^2_{T_2I_2}).$$

Only $E(I_1I_2)$ must be evaluated from the linear functions of I_1 and I_2. If, for example, in the simplest case where X_1 and X_2 are the records for trait 1 and 2 on the animal being evaluated, $I_1 = b_{11}X_1 + b_{12}X_2$ and $I_2 = b_{21}X_1 + b_{22}X_2$, then:

$$E(I_1I_2) = b_{11}b_{21}\sigma^2_{X_1} + (b_{11}b_{22} + b_{12}b_{21})\sigma_{X_1X_2} + b_{12}b_{22}\sigma^2_{X_2}\ .$$

CHAPTER 21

RESTRICTED SELECTION INDEX

Sometimes one trait is at an optimum level (when an intermediate is desirable) but is correlated with another trait of economic importance. Ordinary selection for the economic trait would lead to an unwanted correlated response in the trait which is at an optimum level.

EQUATIONS TO FIND WEIGHTS WITH RESTRICTION

The general problem is to maximize $T = v_1G_1 + v_2G_2 + \cdots + v_mG_m$ but at the same time force the N - m other traits not to change from their present genetic level, i.e., $\Delta G_{m+1} = 0 = \Delta G_{m+2} = \cdots = \Delta G_N$. A solution to this problem is the restricted selection index given by Kempthorne and Nordskog (1959).

In the simplest case $T = v_1G_1$ and the restriction is to be $\Delta G_2 = 0$. Available are measures on the two traits, X_1 and X_2. Selection for $T = v_1G_1$ is by $I^* = b_1^* X_1 + b_2^* X_2$ where the * indicates the restricted selection index; restricted in that the index is to maximize ΔT with the restriction that $\Delta G_2 = 0$.

The restriction, $\Delta G_2 = 0$, is equivalent to the equation for correlated response $\frac{\text{Cov}(G_2,I^*)}{\sigma_{I^*}} D = 0$ so that $\text{Cov}(G_2,I^*) = b_1^* \sigma_{X_1G_2} + b_2^* \sigma_{X_2G_2}$ must be zero and is the restriction.

In addition, the equations for the b's to maximize r_{TI^*} are:

$$\sigma^2_{X_1} b^*_1 + \sigma_{X_1 X_2} b^*_2 = \sigma_{X_1 T}$$
$$\sigma_{X_1 X_2} b^*_1 + \sigma^2_{X_2} b^*_2 = \sigma_{X_2 T}$$

Thus there are three equations including the restriction but only two unknowns. To find a solution a dummy unknown is added; the so-called LaGrange multiplier, λ. The three equations in three unknowns can now be solved. The coefficient matrix is symmetrical as before:

$$b^*_1 \sigma^2_{X_1} + b^*_2 \sigma_{X_1 X_2} + \lambda \sigma_{X_1 G_2} = \sigma_{X_1 T}$$
$$b^*_1 \sigma_{X_1 X_2} + b^*_2 \sigma^2_{X_2} + \lambda \sigma_{X_2 G_2} = \sigma_{X_2 T}$$
$$b^*_1 \sigma_{X_1 G_2} + b^*_2 \sigma_{X_2 G_2} + 0 = 0 \quad .$$

The restricted index will be $I^* = b^*_1 X_1 + b^*_2 X_2$. A solution for λ can be obtained but is not needed for the index.

These equations can be derived by minimizing $E[(T-I^*)^2]$ with the restriction that $2\lambda (b^*_1 \sigma_{X_1 G_2} + b^*_2 \sigma_{X_2 G_2}) = 0$, i.e., equate to zero the partial derivatives of $\sigma^2_T + \sigma^2_{I^*} - 2\sigma_{TI^*} + 2\lambda\sigma_{G_2 I^*}$ with respect to b^*_1, b^*_2, and λ.

If selection is for more than one trait with restriction of more than one trait the procedure can be expanded, instead of a single λ there will be λ_i ($i = m+1, ..., N$) where N - m is the number of traits to hold constant and N is the number of economic traits.

As an example consider m = 2, with $T = v_1 G_1 + v_2 G_2$, and N - m = 2 with the restrictions: $\Delta G_3 = 0 = \Delta G_4$.

The restricted index will be $I^* = \beta^*_1 X_1 + \beta^*_2 X_2 + \beta^*_3 X_3 + \beta^*_4 X_4$.

The restriction equations are:

$$\sigma_{G_3 I^*} = \beta_1^* \sigma_{X_1 G_3} + \beta_2^* \sigma_{X_2 G_3} + \beta_3^* \sigma_{X_3 G_3} + \beta_4^* \sigma_{X_4 G_3} = 0 \text{ and}$$

$$\sigma_{G_4 I^*} = \beta_1^* \sigma_{X_1 G_4} + \beta_2^* \sigma_{X_2 G_4} + \beta_3^* \sigma_{X_3 G_4} + \beta_4^* \sigma_{X_4 G_4} = 0 .$$

Thus, λ_1 and λ_2 will be the LaGrange multipliers in the equations to find the restricted selection index weights:

$$\begin{aligned}
&\beta_1^* \sigma^2_{X_1} + \beta_2^* \sigma_{X_1 X_2} + \beta_3^* \sigma_{X_1 X_3} + \beta_4^* \sigma_{X_1 X_4} + \lambda_1 \sigma_{X_1 G_3} + \lambda_2 \sigma_{X_1 G_4} = \sigma_{X_1 T} \\
&\beta_1^* \sigma_{X_1 X_2} + \beta_2^* \sigma^2_{X_2} + \beta_3^* \sigma_{X_2 X_3} + \beta_4^* \sigma_{X_2 X_4} + \lambda_1 \sigma_{X_2 G_3} + \lambda_2 \sigma_{X_2 G_4} = \sigma_{X_2 T} \\
&\beta_1^* \sigma_{X_1 X_3} + \beta_2^* \sigma_{X_2 X_3} + \beta_3^* \sigma^2_{X_3} + \beta_4^* \sigma_{X_3 X_4} + \lambda_1 \sigma_{X_3 G_3} + \lambda_2 \sigma_{X_3 G_4} = \sigma_{X_3 T} \\
&\beta_1^* \sigma_{X_1 X_4} + \beta_2^* \sigma_{X_2 X_4} + \beta_3^* \sigma_{X_3 X_4} + \beta_4^* \sigma^2_{X_4} + \lambda_1 \sigma_{X_4 G_3} + \lambda_2 \sigma_{X_4 G_4} = \sigma_{X_4 T} \\
&\beta_1^* \sigma_{X_1 G_3} + \beta_2^* \sigma_{X_2 G_3} + \beta_3^* \sigma_{X_3 G_3} + \beta_4^* \sigma_{X_4 G_3} + 0 + 0 = 0 \\
&\beta_1^* \sigma_{X_1 G_4} + \beta_2^* \sigma_{X_2 G_4} + \beta_3^* \sigma_{X_3 G_4} + \beta_4^* \sigma_{X_4 G_4} + 0 + 0 = 0
\end{aligned}$$

EXPECTED RESPONSES

The expected response in selecting according to I^* should be compared to the response in selecting directly for $T = v_1 G_1 + v_2 G_2$ by $I = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$ with no restriction on change in traits 3 and 4. Comparison could also be made with selection for T using just X_1 and X_2. Although G_3 and G_4 may be optimum, the restriction to maintain that optimum may be so costly in terms of ΔG_1 and ΔG_2 that a better procedure would be to let G_3 and G_4 change while selecting strongly for T. A look at the correlated responses may help to answer that question.

In the typical example total response:

$$\Delta T = \frac{Cov(T, I^*)}{\sigma_{I^*}} D = v_1 \Delta G_1 + v_2 \Delta G_2$$

where as usual $\Delta G_1 = \frac{Cov(G_1, I^*)}{\sigma_{I^*}} D$ and $\Delta G_2 = \frac{Cov(G_2, I^*)}{\sigma_{I^*}} D$.

Because of the restriction, $\Delta G_3 = \Delta G_4 = 0$. These responses would be compared with $v_1 \Delta G_1 + v_2 \Delta G_2 + v_3 \Delta G_3 + v_4 \Delta G_4$ as calculated from using the unrestricted index where v_3 and v_4 are the losses in changing G_3 and G_4 from their present optimums. The losses in traits 3 and 4 may be different when the changes are negative from when the changes are positive, i.e., the economic values of traits 3 and 4 may not be linear.

EXAMPLES USING THE RESTRICTED SELECTION INDEX

Records On Animal Being Evaluated

Assume trait 1, is to be improved and trait 2, is to be held constant. A record on each trait is available on all animals to be evaluated.

Given: $\sigma^2_{P_1} = (2500 \text{ lb})^2$ $\sigma^2_{P_2} = (.3\%)^2$ $\sigma_{P_1 P_2} = -150$

$\sigma^2_{G_1} = (1250 \text{ lb})^2$ $\sigma^2_{G_2} = (.21\%)^2$ $\sigma_{G_1 G_2} = -157.5$

$r_{P_{12}} = -.2$, $r_{g_{12}} = -.6$, $h^2_1 = .25$, $h^2_2 = .49$.

Let $v_1 = 1$ since selection is to improve only one trait.

Find: $I^*_1 = b^*_1 X_1 + b^*_2 X_2$.

The equations determining the weights are:

$$(2500)^2\, b_1^* \quad -150\, b_2^* \quad -157.5\lambda \quad = (1250)^2$$

$$-150\, b_1^* + (.3)^2\, b_2^* + (.21)^2\lambda = -157.5$$

$$-157.5\, b_1^* + (.21)^2\, b_2^* \qquad\qquad = 0 \quad \text{so that}$$

$I_1^* = .1581\, X_1 + 564.52\, X_2$,

$\sigma^2_{I_1^*} = (.1581)^2\,(2500)^2 + (564.52)^2\,(.3)^2 + 2(.1581)(564.52)(-150) = 158{,}065$, and

$\sigma_{I^*} = 397.57$.

$$\Delta G_1 = \frac{\text{Cov}(G_1, I_1^*)}{\sigma_{I_1^*}} D = \frac{.1581(1250)^2 + 564.52(-157.5)}{397.57} D = 397.57\ D\ \text{(lb)}.$$

$$\Delta G_2 = \frac{\text{Cov}(G_2, I_1^*)}{\sigma_{I_1^*}} D = \frac{.1581(-157.50) + 564.52(.21)^2}{397.57} D = 0.$$

If selection is for G_1 with no restriction on G_2:

$I_1 = .2167\, X_1 - 1388.89\, X_2$ with $\sigma_{I_1} = 746.5$, so that;

$\Delta G_1 = 746.5\ D$ (lb) and

$$\Delta G_2 = \frac{\text{Cov}(G_2, I_1)}{\sigma_{I_1}} D = \frac{.2167(-157.5) - 1388.89(.21)^2}{746.5} D = -.1278\ D\ (\%).$$

If selection is for G_1 from X_1 only:

$I_1 = .25X_1$ with $\sigma_{I_1} = 625$, so that;

$\Delta G_1 = 625\ D$ (lb), and $\Delta G_2 = \dfrac{.25(-157.5)}{625} D = -.0630\ D\ (\%)$.

The expected responses for the indexes should be compared for their economic impact to determine if restriction is what is really wanted.

Records On Paternal Half-sib Progeny Of Sire Being Evaluated

Evaluate sires by 3 procedures based on

1) $I_1^* = b_1^* X_{11} + b_2^* X_{12}$, 2) $I_1 = b_1 X_{11} + b_2 X_{12}$, and 3) $I_1 = b_1 X_{11}$

where X_{11} is the progeny average for trait 1, X_{12} is the progeny average for trait 2 with 1 record per progeny and p = 20 progeny. I* for 1) is to improve G_1 and not change G_2. I for 2) is to maximize ΔG_1 and I for 3) is to maximize ΔG_1.

Given: σ_{P_1} = 2500 lb, σ_{P_2} = .3%, h_1^2 = .25, h_2^2 = .49, $r_{g_{12}}$ = -.6, $r_{p_{12}}$ = -.2.

Find the indexes and expected response in trait 1 for all 3 procedures.

1) Restricted index:

$I^* = b_1^* X_{11} + b_2^* X_{12}$ and $\text{Cov}(G_{\alpha 2}, I^*) = b_1^* a_{1\alpha} \sigma_{G_2 G_1} + b_2^* a_{1\alpha} \sigma^2_{G_2} = 0$

Equations:

$$\left(\frac{1+(p-1).25\, h_1^2}{p}\right) \sigma^2_{P_1}\, b_1^* + \frac{\sigma_{P_1 P_2} + (p-1).25\, \sigma_{G_1 G_2}}{p}\, b_2^* + .5\, \sigma_{G_1 G_2} \lambda = .5\, \sigma^2_{G_1}$$

$$\frac{\sigma_{P_1 P_2} + (p-1).25\, \sigma_{G_1 G_2}}{p}\, b_1^* + \left(\frac{1+(p-1).25\, h_2^2}{p}\right) \sigma^2_{P_2}\, b_2^* + .5\, \sigma^2_{G_2} \lambda = .5\, \sigma_{G_1 G_2}$$

$$.5\sigma_{G_1 G_2}\, b_1^* + \qquad .5\sigma^2_{G_2}\, b_2^* + \qquad 0\ \lambda = 0$$

Numerically:

$$683{,}594\ b_1^* \qquad -45\ b_2^* - 78.75\lambda = 781{,}250$$

$$-45\ b_1^* + .01497375\ b_2^* + .02205\lambda = -78.75$$

$$-78.75\ b_1^* + .02205\ b_2^* \qquad = 0 \quad \text{so that}$$

$I^* = .9039\ X_{11} + 3228\ X_{12}$, and $\sigma^2_{I^*} = 451{,}951$ with $\sigma_{I^*} = 672.3$.

$$\Delta G_1 = \frac{\text{Cov}(G_{\alpha 1},\ I^*)}{\sigma_{I^*}}\ D = \frac{.9039(.5\ \sigma^2_{G_1}) + 3228(.5\ \sigma_{G_1 G_2})}{672.3}\ D = 672.3\ D \text{ (lb)}$$

$$\Delta G_2 = \frac{\text{Cov}(G_{\alpha 2},\ I^*)}{\sigma_{I^*}}\ D = \frac{.9039(.5\ \sigma_{G_1 G_2}) + 3228(.5\ \sigma^2_{G_2})}{672.3}\ D = 0\ (\%)$$

2) Unrestricted index:

$I = b_1\ X_{11} + b_2\ X_{12}$: equations are upper 2x2 for 1) with same 2 right-hand sides.

Then:

$I = .9931\ X_{11} - 2275\ X_{12}$ and $\sigma^2_I = 955{,}003$ with $\sigma_I = 977.2$, so that

$$\Delta G_1 = \frac{\text{Cov}(G_{\alpha 1},\ I)}{\sigma_I}\ D = \frac{.9931(.5\ \sigma^2_{G_1}) - 2275(.5\ \sigma_{G_1 G_2})}{977.2}\ D = 977.2\ D \text{ (lb)}$$

$$\Delta G_2 = \frac{\text{Cov}(G_{\alpha 2},\ I)}{\sigma_I}\ D = \frac{.9931(.5\ \sigma_{G_1 G_2}) - 2275(.5\ \sigma^2_{G_2})}{977.2}\ D = -.1314\ D\ (\%)$$

3) Single trait index:

$I = b_1 X_{11}$: equation is first diagonal and first RHS of 1).

$I = 1.1429\, X_{11}$ and $\sigma_I^2 = 892{,}857$ with $\sigma_I = 944.9$, so that

$$\Delta G_1 = \frac{\text{Cov}(G_{\alpha 1}, I)}{\sigma_I} D = \frac{1.1429(.5\, \sigma^2_{G_1})}{944.9} D = 944.9\ D\ \text{(lb)}$$

$$\Delta G_2 = \frac{\text{Cov}(G_{\alpha 2}, I)}{\sigma_I} D = \frac{1.1429(.5\, \sigma_{G_1 G_2})}{944.9} D = -.0952\ D\ (\%)$$

SUMMARY OF EXAMPLE

	Female selection		Male selection	
Procedure	$\Delta G_1/D$	$\Delta G_2/D$	$\Delta G_1/D$	$\Delta G_2/D$
$I_1^* = b_1^* X_{11} + b_2^* X_{12}$	398 lb	0 %	672 lb	0 %
$I_1 = b_1 X_{11} + b_2 X_{12}$	747 lb	-.128 %	977 lb	-.131 %
$I_1 = b_1 X_{11}$	625 lb	-.063 %	945 lb	-.095 %

OTHER RESTRICTIONS

The theory and application of restricted selection indexes has been extended to cases other than forcing expected change in certain traits to be zero. These restrictions include directional restriction, proportional change, and specified change other than zero. These restrictions are somewhat more complicated to apply than the zero change restriction and are not be discussed here.

CHAPTER 22

INDEX AND ECONOMIC VALUES IN RETROSPECT

The index in retrospect is an index that might have been used for selection to produce gains that have occurred even though the weights for the index might have been unknown at the time of selection (Dickerson et al., 1954). Determining the index that might have been used depends on finding an index which would have given the set of phenotypic selection differentials actually observed.

Let $I = \sum_{i=1}^{N} w_i P_i$ be the underlying but unknown index that might have been used for selection and D be the selection intensity factor. The phenotypic record for trait i measured on the animal being selected is P_i.

INDEX IN RETROSPECT FROM PHENOTYPIC SELECTION DIFFERENTIALS

If the underlying unknown index is I, the regression of P_j on I gives the expected phenotypic selection response (differential) for trait j, (j = 1, ... , N):

$$\Delta P_j = \frac{Cov(P_j, I)}{\sigma_I^2} \Delta I = \frac{Cov(P_j, I)}{\sigma_I} D = (D/\sigma_I)\ Cov(P_j, I).$$

Because D/σ_I is a constant for all traits, the proportionality of the right-hand sides for different traits will not change. Both D and I, however, may be different for males and

females and even from generation to generation. Indexes in retrospect can be computed separately for males and females and for each generation. If D/σ_I is set to one, then the expectations of the N phenotypic selection differentials are:

$$\begin{aligned}
E[\Delta P_1] &= \mathrm{Cov}(P_1,I) = w_1\sigma^2_{P_1} + w_2\sigma_{P_1P_2} + \cdots + w_N\sigma_{P_1P_N} \\
E[\Delta P_2] &= \mathrm{Cov}(P_2,I) = w_1\sigma_{P_1P_2} + w_2\sigma^2_{P_2} + \cdots + w_N\sigma_{P_2P_N} \\
&\vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \\
E[\Delta P_N] &= \mathrm{Cov}(P_N,I) = w_1\sigma_{P_1P_N} + w_2\sigma_{P_2P_N} + \cdots + w_N\sigma^2_{P_N} \quad .
\end{aligned}$$

Note that the coefficients of the **w**'s are the same as for finding the best selection index weights, i.e., the phenotypic variances and covariances. The selection differentials can be equated to their expectations, i.e., to the left hand sides of the usual selection index equations, to determine in retrospect the relative weights for the index. Here the symbol **w** is used for the weights for the index in retrospect. The phenotypic variances and covariances must be known as well as the phenotypic selection differentials. A linear index in the phenotypic values is assumed as is truncation selection based on the underlying but unknown index.

In matrix notation $\mathbf{w} = \mathbf{P}^{-1}\Delta\mathbf{p}$ where **w** is the vector of retrospective weights, P is the phenotypic variance-covariance matrix, and $\Delta\mathbf{p}$ is the vector of phenotypic selection differentials.

The relative expected correlated responses from using the retrospective index are as before: $\Delta G_j = (D/\sigma_I)\,\mathrm{Cov}(G_j,I)$ $(j=1, \ldots, N)$ which could be compared with the expected responses from the theoretically best index for which economic values are assumed known.

INDEX IN RETROSPECT FROM GENETIC SELECTION DIFFERENTIALS

Another approach for finding the index in retrospect depends on knowing the genetic selection differentials, ΔG_j (j=1, ... , N). Usually each genetic selection differential would be estimated as the difference in phenotypic means between animals of two generations. If breeding values can be estimated for all animals the genetic selection differential can be calculated from the difference in averages of estimated breeding values for the two generations. The underlying I and also D may be different for males and females which may cause a problem in assigning the fractions of ΔG due to male and female selection.

Again let $I = \Sigma w_i P_i$ be the underlying index. The regression of G_j on I will give the expected genetic selection differential for trait j:

$$E[\Delta G_j] = \frac{Cov(G_j, I)}{\sigma_I^2} \Delta I = \frac{Cov(G_j, I)}{\sigma_I} D = (D/\sigma_I)\, Cov(G_j, I) \quad .$$

If D/σ_I is set to 1, the expected values of the genetic selection differentials are:

$$\begin{aligned}
E[\Delta G_1] &= Cov(G_1,I) &&= w_1\sigma^2_{G_1} &&+ w_2\sigma_{G_1G_2} &&+ \cdots + w_N\sigma_{G_1G_N} \\
E[\Delta G_2] &= Cov(G_2,I) &&= w_1\sigma_{G_1G_2} &&+ w_2\sigma^2_{G_2} &&+ \cdots + w_N\sigma_{G_2G_N} \\
&\vdots &&\vdots &&\vdots &&\vdots \\
E[\Delta G_N] &= Cov(G_N,I) &&= w_1\sigma_{G_1G_N} &&+ w_2\sigma_{G_2G_N} &&+ \cdots + w_N\sigma^2_{G_N} \quad .
\end{aligned}$$

Thus, if the genetic variance-covariance matrix is known as well as the genetic selection differentials, the weights for the underlying index can be estimated by equating the estimated genetic selection differentials to the right hand sides of the above equations as $w = G^{-1}\Delta g$. In most cases, however, the phenotypic variance-covariance matrix is much easier to estimate accurately than the genetic variance-covariance matrix, G.

ECONOMIC VALUES DETERMINED FOR THE INDEX IN RETROSPECT

After the retrospective index $I = \Sigma w_i P_i$ is determined, the relative economic weights in retrospect can also be determined if the assumption is true that the retrospective index is the best index for some retrospective economic true value, $T = \Sigma v_i G_i$. Thus, the usual equations to find the selection index weights (which now have estimates) can be used to find corresponding economic values. The calculated numerical values on the left-hand sides which depend on the phenotypic variance-covariance matrix are equated to the right-hand sides which are a function of the economic values to be solved for and the genetic variance-covariance matrix:

$$\sigma^2_{P_1} w_1 + \sigma_{P_1 P_2} w_2 + \cdots + \sigma_{P_1 P_N} w_N = v_1 \sigma^2_{G_1} + v_2 \sigma_{G_1 G_2} + \cdots + v_N \sigma_{G_1 G_N}$$

$$\sigma_{P_1 P_2} w_1 + \sigma^2_{P_2} w_2 + \cdots + \sigma_{P_2 P_N} w_N = v_1 \sigma_{G_2 G_1} + v_2 \sigma^2_{G_2} + \cdots + v_N \sigma_{G_2 G_N}$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$\sigma_{P_1 P_N} w_1 + \sigma_{P_2 P_N} w_2 + \cdots + \sigma^2_{P_N} w_N = v_1 \sigma_{G_N G_1} + v_2 \sigma_{G_N G_2} + \cdots + v_N \sigma^2_{G_N}$$

This procedure is equivalent to equating the phenotypic selection differentials to the right-hand sides of the selection index equations for predicting total merit, the $\sigma_{X_i T}$, and then solving for the economic values. Accurate estimates of the genetic variances and covariances are necessary for determining the economic values in retrospect:

$$\Delta P_1 = v_1 \sigma^2_{G_1} + \cdots + v_N \sigma_{G_1 G_N}$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\Delta P_N = v_1 \sigma_{G_N G_1} + \cdots + v_N \sigma^2_{G_N} .$$

In matrix notation $\mathbf{v} = G^{-1}\Delta p$ where $\mathbf{v}$ is the vector of economic values, G is the genetic variance-covariance matrix and Δp is the vector of phenotypic selection differentials.

AN EMPIRICAL SELECTION INDEX

If the net value of each animal can be determined (even with error if errors in determining economic value are uncorrelated with the X's), then an empirical selection index, $I = \Sigma b_i X_i$, can be found from the multiple regression of net value, y, on the phenotypic traits, X_i (i=1, ..., N):

$$\begin{aligned}
\sigma^2_{X_1}\beta_1 &+ \sigma_{X_1X_2}\beta_2 + \cdots + \sigma_{X_1X_N}\beta_N &&= \sigma_{X_1y} \\
&\vdots &&\vdots \\
\sigma_{X_NX_1}\beta_1 &+ \sigma_{X_NX_2}\beta_2 + \cdots + \sigma^2_{X_N}\beta_N &&= \sigma_{X_Ny}
\end{aligned}$$

The variances and covariances can be estimated from the data which includes net value. The phenotypic variances and covariances also could be estimated from a larger sample of data, some of which does not include net value. In fact, the coefficients on the LHS are the same as for the selection index equations if the phenotypic variances and covariances are known exactly. Solving these equations will give the empirical selection index which is an unbiased estimate of the best index to predict overall economic value. The covariances between total economic value and phenotypic measurements can be used as follows to estimate linear economic values for each trait.

ECONOMIC VALUES FROM EMPIRICAL COVARIANCES

A linear model for total net economic value for animal **m** can be written as: $y_m = \Sigma v_j G_{jm} + e_m$ where $\Sigma v_j G_{jm}$ is overall economic value and e_m is a random error of measurement. Thus, if e_m is uncorrelated with the e's of the X's (perhaps a not very reasonable assumption), then equating: $\sigma_{X_i y} = \sum_j v_j \sigma_{G_i G_j}$ for each trait i would give these equations:

$$\begin{aligned} v_1\sigma^2_{G_1} &+ v_2\sigma_{G_1 G_2} + \cdots + v_N\sigma_{G_1 G_N} = \sigma_{X_1 y} \\ &\vdots \\ v_1\sigma_{G_N G_1} &+ v_2\sigma_{G_N G_2} + \cdots + v_N\sigma^2_{G_N} = \sigma_{X_N y} \end{aligned} .$$

If the genetic variances and covariances are known and the $\sigma_{X_i y}$ have been computed, the equations can be solved to find the economic values.

If the empirical selection index weights are unbiased and because the right-hand sides equal $\Sigma v_j \sigma_{G_i G_j}$ for all i, the economic values can also be estimated from these equations because $I = \Sigma \beta_i X_i$ is a retrospective index:

$$\begin{aligned} \sigma^2_{X_1}\beta_1 &+ \cdots + \sigma_{X_1 X_N}\beta_N = v_1\sigma^2_{G_1} + \cdots + v_N\sigma_{G_1 G_N} \\ &\vdots \\ \sigma_{X_N X_1}\beta_1 &+ \cdots + \sigma^2_{X_N}\beta_N = v_1\sigma_{G_N G_1} + \cdots + v_N\sigma^2_{G_N} \end{aligned}$$

If only the v's are unknown, the equations can be solved to find the economic values where the LHS's are $\sigma_{X_i y}$.

PART TWO

INTRODUCTION

TO

MIXED MODEL

PREDICTION

Chapters 23-33 introduce least squares and mixed model equations for prediction of breeding values, transmitting ability, and real producing ability. The correspondence of selection index and mixed model procedures is demonstrated. How to calculate the inverse of the numerator relationship matrix which is needed for best linear unbiased prediction is described for sire and animal models. The mixed model approach is how modern genetic evaluations are done. Properties of mixed model procedures, however, are not as easy to show as analogous selection index properties without a heavy dose of linear model theory and facility with matrix algebra. Therefore this section serves only as a brief introduction to mixed model methods with emphasis on application.

CHAPTER 23

PREDICTION FROM LINEAR MODELS

Selection index procedures described in Part I require the assumption that phenotypic measurements are perfectly adjusted for all nongenetic factors except the random permanent and temporary environmental effects; i.e., $x_i = y_i - \mu$, where y_i is the actual measurement and μ is a symbolic representation of adjustment for all fixed nongenetic factors such as age effects, year effects, and management effects.

In many situations, the adjustments for the fixed factors must be estimated simultaneously with prediction of genetic values. Some adjustments such as for age may be made from estimates obtained from previous sets of data. Effects of other fixed factors, however, may occur as the records are being made, as for example, the effects of year and management, so that prior estimates of those effects are not available to adjust the records.

A procedure is available for such situations which has many of the properties of the selection index. The procedure is the same as the selection index if all fixed factors are known although the two procedures, at first, appear greatly different.

BEST LINEAR UNBIASED PREDICTION

The mixed model procedure was derived by C. R. Henderson about 1948. He generalized and proved its properties after that time (e.g., C. R. Henderson; 1975, 1984).

The procedure results in what is called best linear unbiased prediction (BLUP) of the random effects such as additive genetic value, transmitting ability, and permanent environmental effects. Best is defined as minimizing the variance of prediction error for procedures which are unbiased and use linear functions of the data. Best linear unbiased predictors can be obtained simultaneously with best linear unbiased estimates (BLUE) of fixed factors from solutions to what are often called Henderson's mixed model equations. With the mixed model equations the predictors of, e.g., additive genetic value are automatically adjusted for the fixed factors as well as possible.

Both BLUP and selection index procedures require the assumption that variances and covariances such as genetic and phenotypic variances and covariances are known. The properties in common between BLUP and selection index are:

1) both are unbiased; the selection index is automatically unbiased whereas BLUP solutions are forced to be unbiased,
2) variances of prediction errors are minimized (the basis for obtaining the equations for both BLUP and selection index),
3) the correlation between the prediction and what is predicted, $r_{T\hat{T}}$, is maximized,
4) if the data and T follow a multivariate normal distribution, then the predictions maximize the probability of correct ranking, and
5) the predictions are the same as selection index except that with BLUP the best linear unbiased estimates of fixed effects are used to adjust the records to a G + E basis whereas with the selection index the true values of the fixed effects are used for adjustment.

The mixed model equations are derived after considerable algebra from minimizing

prediction errors squared and sampling variances of estimates of fixed effects with the condition that the predictions are unbiased. Variances and covariances among the records are considered in an optimum way. The procedure will be illustrated for a few models and will not be covered in general. A complete discussion would require knowledge of matrix algebra and several semesters of statistics (see Henderson, 1984).

MIXED MODEL EQUATIONS

When all observations have the same variance, the mixed model procedure simplifies to a simple set of equations involving all effects in the model except for the residual effects. The number of equations is the same as the number of effects in the model. The procedure is considerably more complex with multiple traits with different variances and covariances. Multiple trait applications will not be discussed.

The equations are the same as ordinary least squares equations if all effects (except residual terms) are fixed effects. The equations are called mixed model equations when random effects or when both random and fixed effects are in the model. The mixed model equations are obtained from simple modifications of the least squares equations. Effects are random if they come from a distribution with some variance such as would be the case for genetic values and real producing abilities. Fixed effects have no variance and theoretically can be repeated exactly. A wide range of effects combine some of the characteristics of both random and fixed effects.

Rules for setting up the mixed model equations will be given for models where each effect in the model is a whole effect (i.e., g_i not $g_i/2$ or a covariate). How to modify these simple rules can be found in most books on applied linear models or statistical methods.

RULES FOR WRITING MIXED MODEL EQUATIONS

1. Compute a sum for each effect in the model excluding residual effects such that each observation that contains the effect is included in the sum.
2. Equate each sum to its model considering all effects as fixed excluding the residual effects. The result is called the ordinary least squares equations (LSE). Put a hat, (^), on the effects to denote solutions to the equations and not actual effects.
3. If an effect comes from a distribution of independent effects with variance, σ^2_v, then add the ratio, σ^2_w/σ^2_v, to the diagonal coefficient of those equations where σ^2_w is the variance of residual effects. Models where the random effects are correlated, e.g., genetic values when animals are related, will be considered by example.
4. Constraints often must be imposed on the equations for fixed effects. The usual rule of thumb is that one nonestimable constraint is needed for all except one classification of fixed effects, e.g., if one constraint is on $\hat{\mu}$ then one classification of fixed effects should not have a constraint imposed. Typical constraints are $\hat{\mu} = 0$ if there is only one fixed classification; $\hat{\mu} = 0$ and the last level of each classification except for one classification also set equal to 0 if there is more than one fixed classification.

INTERPRETATION OF SOLUTIONS

1. Solutions for fixed effects are best linear unbiased estimates (BLUE) of estimable functions of the fixed effects. The jargon concerns interpretation 2.

2. The expected values of the solutions corresponding to fixed effects for models without interaction terms usually have the properties:
   a) E[solution for a fixed effect] $\neq$ actual fixed effect,
   b) E[solutions for fixed effects] depend on the constraints imposed to obtain solutions, and usually
   c) $E\begin{pmatrix}\text{difference in solutions for two fixed} \\ \text{effects in the same classification}\end{pmatrix} = \begin{matrix}\text{difference in the} \\ \text{actual fixed effects}\end{matrix}$
3. Solutions for effects randomly drawn from some distribution of effects such as genetic values are best linear unbiased predictors (BLUP) and have the selection index properties except that the observations have been adjusted for fixed effects with best linear unbiased estimates of the fixed effects rather than by actual values of the fixed effects.

DOT NOTATION

Before the first example in Chapter 24, the dot notation will be introduced which makes writing the equations in a symbolic form less laborious. A more complete discussion of dot notation is in Chapter 34. A dot (period) in place of a subscript signifies summation has occurred over that subscript. Suppose observations are denoted symbolically as P_{ij} where the **i** subscript refers to animal **i** and the **j** subscript refers to the jth record of the animal. Let n_i be the number of records of animal **i**. As an example, let i = 1, 2, or 3, and $n_1 = 2$, $n_2 = 1$, and $n_3 = 4$. Total number of records is:

$$\sum_{i=1}^{3} n_i = n_1 + n_2 + n_3 = n .$$

Similarly, the sum of all records of animal 1 is: $P_{1.} = \sum_{j=1}^{n_1=2} P_{1j} = P_{11} + P_{12}$.

The sum of all records is:

$$P_{..} = P_{1.} + P_{2.} + P_{3.} = \sum_{i=1}^{3} \sum_{j=1}^{n_i} P_{ij} = P_{11} + P_{12} + P_{21} + P_{31} + P_{32} + P_{33} + P_{34}$$

CHAPTER 24

LEAST SQUARES EQUATIONS: ONE-WAY CLASSIFICATION MODEL

The easiest way to describe ordinary least squares and mixed model equations is by example. The simplest example that illustrates least squares equations is the one-way classification model where the classification (sometimes called, factor) describes levels of some type of fixed effect. Suppose records are classified by the age of the animal when the record is made and that each animal has only one record. Then a model is:

$$y_{ij} = \mu + A_i + w_{ij}$$

where μ is a constant,

A_i is the fixed effect of the ith age, and

w_{ij} is the random residual term associated with the record of the jth animal made at the ith age.

Note that a record will always contain the G + E terms, whether stated or not. In this case, $w_{ij} = G_{ij} + E_{ij}$ with two subscripts identifying the animal since the numbering of animals (j) begins at one for each age group (i).

Further suppose the following records are available where the records are equated to their models to clarify rules 1) and 2):

$115 = y_{11} = \mu + A_1 + w_{11}$ $\quad$ $95 = y_{21} = \mu + A_2 + w_{21}$

$85 = y_{12} = \mu + A_1 + w_{12}$ $\quad$ $90 = y_{31} = \mu + A_3 + w_{31}$

$105 = y_{13} = \mu + A_1 + w_{13}$ $\quad$ $110 = y_{32} + \mu + A_3 + w_{32}$

Thus, i = 1, 2, or 3, the number of levels of the fixed factor, age; and $n_1 = 3$, the number of records with age effect 1; $n_2 = 1$; and $n_3 = 2$.

RULE ONE: SUMS

Rule 1 is that a sum is computed for each effect in the model excluding the w_{ij} terms. The four effects in the model are μ, A_1, A_2, and A_3 although any record will contain only two terms; the μ constant and one A_i effect. The sum for μ includes each record having μ in its model which is true for all records. Thus, the sum for the μ equation is $y_{..} = 600$. The sum for the A_1 equation includes each record containing A_1 which is the case for the n_1 records with subscript i = 1. Thus the sum for the A_1 equation is: $y_{1.} = 115 + 85 + 105 = 305$. Similarly the sum for the A_2 equation is $y_{2.} = 95$ and for the A_3 equation is $y_{3.} = 90 + 110 = 200$.

RULE TWO: MODELS OF SUMS

Rule 2 is to equate each sum to its model, excluding the w_{ij} terms. The model for $y_{..}$ is simply the sum of the models for all of the $n_{.}$ records included in the sum:

n_1 of the records have model $\mu + A_1$,

n_2 of the records have model $\mu + A_2$, and

n_3 of the records have model $\mu + A_3$,

so that the model for $y_{..}$ is $(n_1 + n_2 + n_3)\mu + n_1A_1 + n_2A_2 + n_3A_3$. Similarly the model

for $y_{1.}$ is the sum of the models for the n_1 records included in $y_{1.}$ where all n_1 records contain μ and A_1 so the model for $y_{1.}$ is $n_1\mu + n_1A_1$.

The same pattern applies to the model for $y_{2.}$ where all records contain μ and A_2 so that the model, is $n_2\mu + n_2A_2$ and for $y_{3.}$ the model is $n_3\mu + n_3A_3$.

Written in the usual symbolic form and with ^'s to indicate solutions and not necessarily estimates of the effects:

$$\begin{aligned}
\mu: &\quad n_{..}\hat{\mu} + n_1\hat{A}_1 + n_2\hat{A}_2 + n_3\hat{A}_3 &&= y_{..}\\
A_1: &\quad n_1\hat{\mu} + n_1\hat{A}_1 &&= y_{1.}\\
A_2: &\quad n_2\hat{\mu} + n_2\hat{A}_2 &&= y_{2.}\\
A_3: &\quad n_3\hat{\mu} + n_3\hat{A}_3 &&= y_{3.}
\end{aligned}$$

For the example, the equations in numerical form are:

$$\begin{aligned}
6\hat{\mu} + 3\hat{A}_1 + 1\hat{A}_2 + 2\hat{A}_3 &= 600\\
3\hat{\mu} + 3\hat{A}_1 &= 305\\
1\hat{\mu} + 1\hat{A}_2 &= 95\\
2\hat{\mu} + 2\hat{A}_3 &= 200
\end{aligned}$$

Note: 1) The numerical coefficients are symmetrical; i. e., the coefficients in the first row are the same as the coefficients in the first column on the LHS's. The same is true for rows and columns two and three.

2) The off-diagonal coefficients among the A equations are zero because, for example, a record made at age 1 cannot also be made at age 2 or age 3.

RULE THREE: RANDOM EFFECTS

Rule 3 does not apply in this example because the model does not include any random factors other than the residuals, the w_{ij}'s.

RULE FOUR: CONSTRAINTS AND SOLUTIONS

The three "A" equations sum to the "μ" equation. Thus, even though there are four equations in four unknowns, four solutions cannot be obtained because the equations are not independent. To obtain a set of solutions, one constraint must be imposed on the original four solutions.

a) The constraint $\hat{\mu} = 0$ is the easiest constraint to use for the computations. The equation for μ as well as $\hat{\mu}$ is eliminated to maintain symmetry in the remaining equations which become:

$$\begin{aligned} n_1\hat{A}_1 &= y_{1.} \\ n_2\hat{A}_2 &= y_{2.} \\ n_3\hat{A}_3 &= y_{3.} \end{aligned}$$

b) Another possible constraint is to set $\hat{A}_3 = 0$; in that case the equation for A_3 is eliminated as well as $\hat{A}_3$ to maintain symmetry in the remaining equations (i. e., $\hat{A}_3$ vanishes as well as the equation for $y_{3.}$):

$$\begin{aligned} n_{..}\hat{\mu} + n_1\hat{A}_1 + n_2\hat{A}_2 &= y_{..} \\ n_1\hat{\mu} + n_1\hat{A}_1 &= y_{1.} \\ n_2\hat{\mu} + n_2\hat{A}_2 &= y_{2.} \end{aligned}$$

c) A more complex constraint is to set $\hat{A}_1 + \hat{A}_2 + \hat{A}_3 = 0$. This equation is in addition to the least squares equations and to make the numerical coefficients of the equations symmetrical a dummy unknown (called in statistical jargon, a Lagrange multiplier, λ) is added to each equation so that:

$$
\begin{aligned}
n_{\cdot}\hat{\mu} + n_1\hat{A}_1 + n_2\hat{A}_2 + n_3\hat{A}_3 + 0\lambda &= y_{\cdot\cdot} \\
n_1\hat{\mu} + n_1\hat{A}_1 + 1\lambda &= y_{1\cdot} \\
n_2\hat{\mu} + n_2\hat{A}_2 + 1\lambda &= y_{2\cdot} \\
n_3\hat{\mu} + n_3\hat{A}_3 + 1\lambda &= y_{3\cdot} \\
0\hat{\mu} + 1\hat{A}_1 + 1\hat{A}_2 + 1\hat{A}_3 + 0\lambda &= 0
\end{aligned}
$$

The solutions with constraint $\hat{\mu} = 0$ are the easiest to use to discuss the principle involved. The solutions, as can be seen from examining the equations when $\hat{\mu} = 0$ and the $y_{\cdot\cdot}$ equation is eliminated, are:

$\hat{\mu} = 0$, $\hat{A}_1 = y_{1\cdot}/n_1$, $\hat{A}_2 = y_{2\cdot}/n_2$, and $\hat{A}_3 = y_{3\cdot}/n_3$. Note that the constraint, $\hat{\mu} = 0$ is one of the solutions.

EXPECTATIONS OF SOLUTIONS

A result of having to impose a constraint is the necessity to be careful in interpretation of the solutions. Obviously in most cases $E[\hat{\mu}] \neq \mu$ because $E[0] = 0$. The $E[\hat{A}_1]$ can be found easily for the one-way classification model with the constraint $\hat{\mu} = 0$. Notice that $E[y_{ij}] = \mu + A_i$ for all subscripts **j**. Thus,

$$
\begin{aligned}
E[\hat{A}_1] = E[y_{1\cdot}/n_1] &= (1/n_1)E[y_{11} + y_{12} + y_{13} + \ldots + y_{1n_1}] \\
&= (1/n_1)[n_1(\mu + A_1)] = \mu + A_1 .
\end{aligned}
$$

Similarly,

$$E[\hat{A}_2] = \mu + A_2 \text{ and } E[\hat{A}_3] = \mu + A_3 .$$

These results show that μ cannot be estimated, and also that none of the A's can be estimated. What can be estimated are functions of $\mu + A_i$, the models for the records.

For example, $A_1 - A_2$ can be estimated by $\hat{A}_1 - \hat{A}_2$ because:

$$E[\hat{A}_1 - \hat{A}_2] = E[\hat{A}_1] - E[\hat{A}_2] = [(\mu + A_1) - (\mu + A_2)] = A_1 - A_2 .$$

The important principle is that solutions obtained using other constraints will have different expectations but that the same estimates of differences can be obtained. For the constraint, $\hat{A}_3 = 0$, $E[\hat{A}_3] = E[0] = 0$. In fact, with that constraint:

$$E[\hat{A}_1] = A_1 - A_3,\ E[\hat{A}_2] = A_2 - A_3, \text{ and } E[\hat{\mu}] = \mu + A_3.$$

As with the $\hat{\mu} = 0$ constraint, the same estimates of differences can be obtained. The estimate of $A_1 - A_3$ is $\hat{A}_1$, the estimate of $A_2 - A_3$ is $\hat{A}_2$, and the estimate of $A_1 - A_2$ is $\hat{A}_1 - \hat{A}_2$. For the sum to zero constraint, $\hat{A}_1 + \hat{A}_2 + \hat{A}_3 = 0$, the expectations of solutions are a little more difficult to find, but the result is that only $\mu + A_i$ and differences among the A's can be estimated. Estimability is defined as the property that a function of the records can be found that has the expected value desired, e.g., $A_1 - A_2$ or $\mu + A_1$.

Finding the expectations of solutions for more complicated models is more difficult and for ease of computation requires some knowledge of matrix algebra. Those techniques are taught in courses in linear models and are beyond the scope of this book.

CHAPTER 25

THE ANIMAL MODEL

THE ANIMAL MODEL WITHOUT REPEATED RECORDS

A simple one-way random classification model results when records are classified by the animal making the record when no fixed classification effects or other random effects are included in the model. If each animal has only one record, each record is assigned to a separate classification. This model can be used to illustrate the similarity between selection index and BLUP. The cases where each animal can have more than one record and where the animals are related will be discussed later.

The model for a record of animal **i** is:

$$y_i = \mu + g_i + w_i$$

where μ is a constant (several fixed factors could be in the model with equations developed as in Chapter 24),

g_i is the effect on the record of the animal's genotype, usually assumed to be additive genetic effects, with $E[g_i] = 0$ and $E[g_i^2] = \sigma_g^2 = h^2\sigma_y^2$, and

w_i is the residual effect of the sum of environmental effects on y_i, with $E[w_i] = 0$ and $E[w_i^2] = \sigma_w^2 = \sigma_E^2 = (1-h^2)\sigma_y^2$.

The mixed model equations are obtained by setting up the least squares equations (same as considering each animal's additive genetic value as a fixed effect) and then adding σ_w^2/σ_g^2 to the diagonal of the coefficent matrix of each animal (additive genetic value) equation.

Note that $\sigma_w^2/\sigma_g^2 = (1\text{-}h^2)\sigma_y^2/h^2\sigma_y^2 = (1\text{-}h^2)/h^2$. Let $\lambda = (1\text{-}h^2)/h^2$.

Because each animal has only one record, the mixed model equations are especially easy to write and are as follows for three animals:

$$\begin{aligned} 3\hat{\mu} + \hat{g}_1 + \hat{g}_2 + \hat{g}_3 &= y_\cdot \\ \hat{\mu} + (1+\lambda)\hat{g}_1 &= y_1 \\ \hat{\mu} + (1+\lambda)\hat{g}_2 &= y_2 \\ \hat{\mu} + (1+\lambda)\hat{g}_3 &= y_3 \end{aligned}$$

The four equations in four unknowns ($\hat{\mu}$, $\hat{g}_1$, $\hat{g}_2$, $\hat{g}_3$) can be solved without imposing a constraint because when λ is added to the diagonal coefficients, the three animal equations do not sum to the μ equation. The solution, $\hat{\mu}$, will be BLUE of μ because for this model $E[\hat{\mu}] = \mu$. The solutions, $\hat{g}_1$, $\hat{g}_2$, and $\hat{g}_3$, will be BLUP and correspond to selection indexes for additive genetic values of animals 1, 2 and 3.

The correspondence to selection index can be shown by examining any of the animal equations (e.g., animal 3):

$$\hat{\mu} + (1+\lambda)\hat{g}_3 = y_3;$$

$$(1+\lambda)\hat{g}_3 = y_3 - \hat{\mu};$$

$$\hat{g}_3 = h^2(y_3 - \hat{\mu})$$

Note that $(1+\lambda) = 1 + (1\text{-}h^2)/h^2 = [h^2 + (1\text{-}h^2)]/h^2 = 1/h^2$

Thus $\hat{g}_3$ is the same as selection index, $I_3 = h^2(y_3 - \mu)$, except that BLUE of μ, $\hat{\mu}$, is used to adjust the animal's record rather than μ. With only three records, as in this example, $\hat{\mu}$ may be poorly estimated.

To show that $\hat{\mu} = \bar{y}_.$, the average of the 3 records, substitute $\hat{g}_i = h^2(y_i - \hat{\mu})$ for $i = 1, 2, 3$ into the first equation and then do some simplification:

$$\begin{aligned} 3\hat{\mu} \quad &+ h^2(y_1 + y_2 + y_3 - 3\hat{\mu}) &&= y_. \\ 3\hat{\mu} \quad &+ h^2(y_. - 3\hat{\mu}) &&= y_. \\ 3\hat{\mu}(1 - h^2) \quad & &&= y_.(1 - h^2) \qquad \text{and thus} \\ & &\hat{\mu} &= y_./3 = \bar{y}_. \ . \end{aligned}$$

ANIMAL MODEL WITH REPEATED RECORDS

Although the records will be classified in only one way, by animal, the effects associated with animal **i** on its record are of two kinds, g_i and p_i, where g_i is the additive genetic value and p_i is the effect of permanent environmental factors which affect each record of the animal. This model with the permanent environmental effect corresponds to the repeatability model introduced in Chapter 8. Again for simplicity assume no fixed effects except μ in the model for y_{ij} the jth record of animal **i**:

$$y_{ij} = \mu + g_i + p_i + w_{ij}$$

where μ is a constant,

g_i is the additive genetic value with $\sigma_g^2 = h^2\sigma_y^2$,

p_i is the permanent environmental effect associated with all records of animal **i** with $\sigma_p^2 = (r - h^2)\sigma_y^2$, and

w_{ij} is the residual effect (temporary environmental effects) associated with the jth record of animal **i** with $\sigma_w^2 = (1 - r)\sigma_y^2$.

Note that $\sigma_g^2 + \sigma_p^2 + \sigma_w^2 = \sigma_y^2$ and that $\sigma_g^2 + \sigma_p^2 = r\sigma_y^2$.
Thus $\sigma_w^2/\sigma_g^2 = (1 - r)/h^2 = \lambda$ and $\sigma_w^2/\sigma_p^2 = (1 - r)/(r - h^2) = \gamma$. The sum of n_i records on animal **i** will be $y_{i.}$. As an example, consider two animals with n_1 and n_2 records. Five equations will be needed corresponding to the five effects; μ, g_1, g_2, p_1, and p_2. The least squares equations will be identical for g_1 and p_1 and for g_2 and p_2 but λ will be added to the diagonal coefficients of the g equations and γ will be added to the diagonal coefficients of the p equations:

$$
\begin{array}{lccccccccc}
\mu: & n_{\cdot}\hat{\mu} & + & n_1\hat{g}_1 & + & n_2\hat{g}_2 & + & n_1\hat{p}_1 & + & n_2\hat{p}_2 = y_{\cdot\cdot} \\
g_1: & n_1\hat{\mu} & + & (n_1 + \lambda)\hat{g}_1 & + & & & n_1\hat{p}_1 & + & = y_{1\cdot} \\
g_2: & n_2\hat{\mu} & + & & & (n_2 + \lambda)\hat{g}_2 & & & & n_2\hat{p}_2 = y_{2\cdot} \\
p_1: & n_1\hat{\mu} & + & n_1\hat{g}_1 & + & & & (n_1 + \gamma)\hat{p}_1 & & = y_{1\cdot} \\
p_2: & n_2\hat{\mu} & + & & & n_2\hat{g}_2 & + & & & (n_2 + \gamma)\hat{p}_2 = y_{2\cdot}
\end{array}
$$

Again, no constraints are needed because the **g** equations do not sum to the μ equation or to the sum of the **p** equations because of the nonzero ratios λ and γ added to the diagonal coefficients. Thus $E[\hat{\mu}] = \mu$. Solutions, $\hat{g}_1$ and $\hat{g}_2$, correspond to selection index predictions of additive genetic values of animals 1 and 2. Similarly, $\hat{g}_i + \hat{p}_i$ estimates producing ability of animal **i** and corresponds to selection index for producing ability. Both correspondences can be shown by examining the mixed model equations. For example, consider the equations for animal 1, the g_1 and p_1 equations. Because the right-hand sides of the two equations are the same, $y_{1.}$, the left-hand sides must also equal each other. Thus, $n_1\hat{\mu} + (n_1 + \lambda)\hat{g}_1 + n_1\hat{p}_1 = n_1\hat{\mu} + n_1\hat{g}_1 + (n_1 + \gamma)\hat{p}_1$.

Terms in $\hat{\mu}$ and n_1 drop out to leave:

$$\lambda\hat{g}_1 = \gamma\hat{p}_1 \text{ so that } \hat{p}_1 = (\lambda/\gamma)\hat{g}_1 = [(r - h^2)/h^2]\hat{g}_1.$$

Now substitute this expression for $\hat{p}_1$ into the g_1 equation:

$$n_1\hat{\mu} + (n_1 + \lambda)\hat{g}_1 + n_1(\lambda/\gamma)\hat{g}_1 = y_{1.}$$

Thus with reordering:

$$[n_1 + \lambda + n_1(\lambda/\gamma)]\hat{g}_1 = y_{1.} - n_1\hat{\mu}$$

Replace λ with $(1 - r)/h^2$, γ with $(1 - r)/(r - h^2)$ and $y_{1.}$ with $n_1\bar{y}_{1.}$ (the average times n_1 is the sum) and with some algebra:

$$\frac{n_1h^2 + 1 - r + n_1r - n_1h^2}{h^2}\hat{g}_1 = n_1(\bar{y}_{1.} - \hat{\mu})$$

and

$$\frac{1 + (n_1 - 1)r}{h^2}\hat{g}_1 = n_1(\bar{y}_{1.} - \hat{\mu}) \text{ so that } \hat{g}_1 = \frac{n_1h^2}{1+(n_1-1)r}[\bar{y}_{1.} - \hat{\mu}],$$

which is the selection index for g_1 for n_1 records on animal 1 with μ replaced by $\hat{\mu}$. Because producing ability is $g_1 + p_1$, add $\hat{g}_1$ and $\hat{p}_1$ to estimate $g_1 + p_1$:

$$\hat{g}_1 + \hat{p}_1 = \frac{n_1h^2}{1+(n_1 - 1)r}[\bar{y}_{1.} - \hat{\mu}] + \left[\frac{r - h^2}{h^2}\right]\frac{n_1h^2}{1+(n_1 - 1)r}[\bar{y}_{1.} - \hat{\mu}] \text{ so that}$$

$$\hat{g}_1 + \hat{p}_1 = \frac{n_1r}{1+(n_1 - 1)r}[\bar{y}_{1.} - \hat{\mu}],$$

which is the selection index for producing ability with $\hat{\mu}$ instead of μ.

CHAPTER 26

SIRE MODELS

ONE-WAY RANDOM CLASSIFICATION SIRE MODEL

The one-way random classification model applies when the data can be classified according to effects which can be thought of as coming randomly from a distribution of effects. For example, the records may be grouped according to the sires of the animals with records.

Suppose the model is:

$$y_{ij} = \mu + s_i + w_{ij}$$

where μ is a constant,

s_i is an effect common to all animals having sire i; (this effect is equivalent to transmitting ability or one-half additive genetic value of the sire because a sample one-half of his genes are transmitted to each of his n_i progeny), $E[s_i] = 0$ and $E(s_i^2) = \sigma_s^2$ = paternal half-sib covariance, $h^2\sigma_y^2/4$, and

w_{ij} is an effect associated with the record of the jth progeny of the ith sire.

Note $s_i + w_{ij} = G_{ij} + E_{ij}$ so that $E(w_{ij}) = 0$ and $\sigma_w^2 = \sigma_y^2 - \sigma_s^2 = \sigma_y^2(1 - h^2/4)$. Thus, $\sigma_w^2/\sigma_s^2 = (1 - h^2/4)/(h^2/4) = (4 - h^2)/h^2$. Let $\lambda = (4 - h^2)/h^2$.

The mixed model equations now are appropriate and are obtained by setting up the least squares equations (same as considering sire effects as fixed effects) and then adding $\sigma_w^2/\sigma_s^2 = \lambda$ to the diagonals of the coefficient matrix corresponding to the sire equations when the sires are assumed to be unrelated. The λ term essentially takes into account the additive relationships among animals with the same sire as does the selection index procedure.

The mixed model equations become (for the case of three sire groups):

$$\begin{aligned}
n_{\cdot}\hat{\mu} + n_1\hat{s}_1 + n_2\hat{s}_2 + n_3\hat{s}_3 &= y_{\cdot\cdot} \\
n_1\hat{\mu} + (n_1 + \lambda)\hat{s}_1 + &= y_{1\cdot} \\
n_2\hat{\mu} + (n_2 + \lambda)\hat{s}_2 &= y_{2\cdot} \\
n_3\hat{\mu} + (n_3 + \lambda)\hat{s}_3 &= y_{3\cdot}
\end{aligned}$$

and if $h^2 = .25$, then $\lambda = (4 - h^2)/h^2 = 15$. Note that σ_y^2 is not required although the ratio σ_w^2/σ_s^2 must be known.

If $n_1 = 11$, $n_2 = 4$, and $n_3 = 15$, the numerical equations except for the sums are:

$$\begin{aligned}
30\hat{\mu} + 11\hat{s}_1 + 4\hat{s}_2 + 15\hat{s}_3 &= y_{\cdot\cdot} \\
11\hat{\mu} + (11 + 15)\hat{s}_1 &= y_{1\cdot} \\
4\hat{\mu} + (4 + 15)\hat{s}_2 &= y_{2\cdot} \\
15\hat{\mu} + (15 + 15)\hat{s}_3 &= y_{3\cdot}
\end{aligned}$$

Because of the extra diagonal terms, $\lambda = 15$, the sire equations do not sum to the μ equation as was also true with the animal models in Chapter 25. The four equations in four unknowns can be solved without imposing a constraint. The solution, $\hat{\mu}$, is BLUE of μ since for this simple model $E[\hat{\mu}] = \mu$. Solutions, $\hat{s}_1$, $\hat{s}_2$, and $\hat{s}_3$, are BLUP and correspond to selection indexes for transmitting ability of sires 1, 2, and 3.

The equivalence of BLUP to selection index can be shown by looking at any of the sire equations, e.g., the equation for sire 1, and noting that $\bar{y}_{1.} = y_{1.}/n_1$, the average of progeny of sire 1:

$$n_1\hat{\mu} + (n_1 + 15)\hat{s}_1 = y_{1.}$$

$$(n_1 + 15)\hat{s}_1 = n_1(\bar{y}_{1.} - \hat{\mu}) .$$

Thus: $$\hat{s}_1 = \left[\frac{n_1}{n_1 + 15}\right](\bar{y}_{1.} - \hat{\mu})$$

which is the same prediction as with the selection index except that the BLUE of μ, $\hat{\mu}$, instead of μ is subtracted from the progeny average to adjust for the fixed constant in the model of each record.

TWO-WAY FIXED AND RANDOM (SIRE) CLASSIFICATION MODEL

Most mixed model analyses are for models that contain both fixed and random effects. The sire model with one fixed factor, such as management effects, is an example.

Assume the model is:

$$y_{ijk} = \mu + m_i + s_j + w_{ijk}$$

where:

- y_{ijk} is the record of progeny k of sire j made in management level i,
- m_i is the fixed effect of management i,
- s_j is an effect common to progeny of sire j with variance σ^2_s, with σ^2_s = paternal half-sib covariance = $h^2\sigma^2_y/4$, and
- w_{ijk} is a random residual effect associated with the record of progeny k of sire j made in management level i, with variance $\sigma^2_w = \sigma^2_y - \sigma^2_s = (1 - h^2/4)\sigma^2_y$.

For example, assume $h^2 = .25$, then $\sigma^2_w/\sigma^2_s = \lambda = 15$.

As an example, assume the following observations have been made:

$y_{111} = 530$	$y_{131} = 350$	$y_{211} = 380$	$y_{221} = 410$
$y_{112} = 520$	$y_{132} = 340$	$y_{212} = 400$	$y_{222} = 440$
$y_{121} = 460$	$y_{133} = 300$	$y_{213} = 410$	

The first subscript (i) denotes the management level and the second subscript (j) identifies the sire of the animal. The largest value of the third subscript (k) for a particular combination of i and j denotes the number of observations for that combination, n_{ij}. For this example: i = 1 or 2; j = 1, 2, or 3; and

$$n_{11} = 2, n_{12} = 1, n_{13} = 3, n_{21} = 3, n_{22} = 2, n_{23} = 0.$$

There are 6 effects in the model: μ, m_1, m_2, s_1, s_2, and s_3.

Equation for μ:

All the observations contain μ so that the sum for μ is $y_{...} = 4540$.

Equation for m_1:

All observations with i = 1 contain m_1 so that the sum for m_1 is $y_{1..} = 2500$.

Equation for m_2:

The sum of observations with i = 2 is $y_{2..} = 2040$.

Equation for s_1:

All observations with j = 1 contain s_1 so that the sum for s_1 is $y_{.1.} = 2240$.

Equation for s_2:

The sum for observations with j = 2 is $y_{.2.} = 1310$.

Equation for s_3:

The sum for observations with j = 3 is $y_{.3.} = 990$.

Usually the easiest way to set up the equations with pencil and paper is to make tables of the subclass numbers and sums:

n_{ij}

	1	2	3	$n_{i.}$
1	2	1	3	6
2	3	2	0	5
$n_{.j}$	5	3	3	11

$y_{ij.}$

	1	2	3	$y_{i..}$
1	1050	460	990	2500
2	1190	850	----	2040
$y_{.j.}$	2240	1310	990	4540

The least squares equations in symbolic form are:

$$\begin{array}{lccccccccccc}
\mu: & n_{..}\hat{\mu} & + & n_{1.}\hat{m}_1 & + & n_{2.}\hat{m}_2 & + & n_{.1}\hat{s}_1 & + & n_{.2}\hat{s}_2 & + & n_{.3}\hat{s}_3 & = y_{...} \\
m_1: & n_{1.}\hat{\mu} & + & n_{1.}\hat{m}_1 & & & + & n_{11}\hat{s}_1 & + & n_{12}\hat{s}_2 & + & n_{13}\hat{s}_3 & = y_{1..} \\
m_2: & n_{2.}\hat{\mu} & + & & + & n_{2.}\hat{m}_2 & + & n_{21}\hat{s}_1 & + & n_{22}\hat{s}_2 & + & n_{23}\hat{s}_3 & = y_{2..} \\
s_1: & n_{.1}\hat{\mu} & + & n_{11}\hat{m}_1 & + & n_{21}\hat{m}_2 & + & n_{.1}\hat{s}_1 & & & & & = y_{.1.} \\
s_2: & n_{.2}\hat{\mu} & + & n_{12}\hat{m}_1 & + & n_{22}\hat{m}_2 & + & & & n_{.2}\hat{s}_2 & & & = y_{.2.} \\
s_3: & n_{.3}\hat{\mu} & & n_{13}\hat{m}_1 & + & n_{23}\hat{m}_2 & + & & & & & n_{.3}\hat{s}_3 & = y_{.3.}
\end{array}$$

The n_{ij} table summarizes the number of each effect in each sum. For example, the sum for m_1 includes $n_{1.}$ records. Each of those records contains μ and m_1. Obviously none contains the m_2 effect. The number of records with m_1 also containing s_1 is n_{11}, n_{12} contain s_2 and n_{13} contain s_3. Note that the first row (i=1) of the n_{ij} table consists of n_{11}, n_{12}, and n_{13}. Similarly the sum for s_1 includes $n_{.1}$ records each containing μ and s_1. The first column (j=1) of the n_{ij} table consists of n_{11} and n_{21}, the number of records containing s_1 which also contain respectively, effects m_1 and m_2.

To convert the least squares equations to mixed model equations, the ratio, $\sigma_w^2/\sigma_s^2 = 15$ for heritability of .25, is added to the diagonal coefficient of each of the s equations so that they become $(n_{.1} + 15)\hat{s}_1$, $(n_{.2} + 15)\hat{s}_2$, and $(n_{.3} + 15)\hat{s}_3$.

In numerical form the mixed model equations are:

$$
\begin{aligned}
11\hat{\mu} + 6\hat{m}_1 + 5\hat{m}_2 + 5\hat{s}_1 + 3\hat{s}_2 + 3\hat{s}_3 &= 4540\\
6\hat{\mu} + 6\hat{m}_1 + 2\hat{s}_1 + 1\hat{s}_2 + 3\hat{s}_3 &= 2500\\
5\hat{\mu} + 5\hat{m}_2 + 3\hat{s}_1 + 2\hat{s}_2 &= 2040\\
5\hat{\mu} + 2\hat{m}_1 + 3\hat{m}_2 + (5+15)\hat{s}_1 &= 2240\\
3\hat{\mu} + 1\hat{m}_1 + 2\hat{m}_2 + (3+15)\hat{s}_2 &= 1310\\
3\hat{\mu} + 3\hat{m}_1 + (3+15)\hat{s}_3 &= 990
\end{aligned}
$$

One constraint must be imposed on either $\hat{\mu}$ or one of the $\hat{m}$'s to obtain a set of solutions. For example, let $\hat{\mu} = 0$ and also eliminate that equation to maintain symmetry of the coefficient matrix. The equations to solve are:

$$
\begin{aligned}
6\hat{m}_1 + 2\hat{s}_1 + \hat{s}_2 + 3\hat{s}_3 &= 2500\\
5\hat{m}_2 + 3\hat{s}_1 + 2\hat{s}_2 &= 2040\\
2\hat{m}_1 + 3\hat{m}_2 + 20\hat{s}_1 &= 2240\\
\hat{m}_1 + 2\hat{m}_2 + 18\hat{s}_2 &= 1310\\
3\hat{m}_1 + 18\hat{s}_3 &= 990
\end{aligned}
$$

Solutions are:

$\hat{\mu}$	= 0	$\hat{s}_1$	= 10
$\hat{m}_1$	= 420	$\hat{s}_2$	= 5
$\hat{m}_2$	= 400	$\hat{s}_3$	= -15

Note that $\hat{s}_1 + \hat{s}_2 + \hat{s}_3 = 0$. This property holds for any classification of random effects that are uncorrelated such as unrelated sires.

The unbiased estimate of $m_1 - m_2$ is $\hat{m}_1 - \hat{m}_2 = 420 - 400 = 20$.

Although how to find expectations of solutions is generally beyond the scope of this book, it is known $E[\hat{\mu}] \neq \mu$, $E[\hat{m}_1] \neq m_1$ and $E[\hat{m}_2] \neq m_2$. Obviously $E[\hat{\mu}_1] = E[0] = 0$. Actually $E[\hat{m}_1] = \mu + m_1$ and $E[\hat{m}_2] = \mu + m_2$ when the constraint $\hat{\mu} = 0$ is used so that $E[\hat{m}_1 - \hat{m}_2] = m_1 - m_2$.

If management levels were considered random effects, then changes in the example would be these:

σ_w^2/σ_m^2 would be added to the diagonals of the management equations,

σ_w^2/σ_s^2 would be added to the diagonals of the sire equations as before,

no constraints would be imposed,

$\hat{m}_1 + \hat{m}_2 = 0$, and

$E[\hat{\mu}] = \mu$ when no other fixed effects are in the model.

In this case $\sigma_y^2 = \sigma_m^2 + \sigma_s^2 + \sigma_w^2$ so that σ_w^2/σ_s^2 may be different from when management levels are considered to be fixed effects.

CHAPTER 27

COMPUTING THE INVERSE OF THE ADDITIVE RELATIONSHIP MATRIX

When genetic values are to be predicted with mixed model procedures the inverse of the additive relationship matrix is used to account for the covariances among the genetic effects among the animals in the model. A logical procedure would seem to be to first calculate the relationship matrix using the tabular method and then have a computer program calculate the inverse. The problem is that even the most powerful computer cannot calculate the inverse for more than 10 to 20 thousand animals in a reasonable amount of time. Henderson (1976) solved the computing problem by finding a rapid way of calculating the inverse of the relationship matrix directly without ever calculating the relationship matrix. If the animals are not inbred, or are assumed not to be inbred, the procedure is very easy. Ignoring a small amount of inbreeding probably is a good approximation in most prediction problems. The exact procedure that accounts for inbreeding is easy with a computer but the explanation is beyond the scope of this book.

The computing steps for each animal with the assumption of no inbreeding involve adding from one to nine values for each animal to different elements of the inverse of the relationship matrix depending on how many parents are known. After all animals have been processed in this way, the result is the inverse of the relationship matrix. Then the inverse elements are multiplied by the proper variance ratio, σ^2_w/σ^2_v, depending on the model, and

are used to modify the least squares equations to construct the mixed model equations; for animal models $\sigma_v^2 = \sigma_g^2$ and for sire models $\sigma_v^2 = \sigma_s^2 = \sigma_g^2/4$ and σ_w^2 is the corresponding residual variance in each case.

The animals can be processed in any order if inbreeding is ignored. Base animals must be included even though they may not have records. Base animals are animals that establish relationships among other animals but are not themselves related.

Because the base animals and some other animals to be evaluated such as sires for sex-limited traits may not have records, the mixed model equations will be augmented as will be illustrated in Chapter 28 to include an equation for each base or other animal without a record with a zero sum on the right-hand-side. The coefficients on the left-hand-sides for those animals will consist only of terms from the inverse of the additive relationship matrix multiplied by σ_w^2/σ_v^2. Because the model for the zero right-hand-side is zero there are no least squares coefficients for the base animal equations.

Any base animal with only one descendent with records (for the sire model only one son with progeny with records) need not be included in the inverse calculation or in the augmented mixed model equations. Such a base parent would be listed as unknown in the calculation of the inverse of the relationship matrix. If such a base animal is included in the inverse of the relationship matrix, then an equation must be included in the mixed model equations for that animal as for any other base animal. Solutions for animals with records will be the same either way if the correct additive relationships are used.

Because the computing procedure can accept animals in any order, putting the base animals at the end of the inverse matrix, rather than the beginning when calculating relationships, may make setting up the equations easier to set up than putting them first or ordering by age.

RULES WHEN SIRES AND DAMS ARE KNOWN

The simple rules for building the inverse of the relationship matrix for non-inbred animals from pedigrees with sires and dams are:

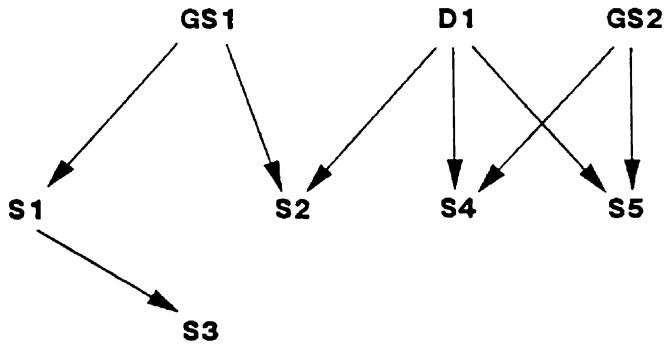
If known			
Animal p	Sire s	Dam d	Then add what to where [a/]
Yes	No	No	1 to (p,p)
Yes	Yes	No	4/3 to (p,p); -2/3 to (s,p); 1/3 to (s,s)
Yes	No	Yes	4/3 to (p,p); -2/3 to (d,p); 1/3 to (d,d)
Yes	Yes	Yes	2 to (p,p); -1 to (s,p) and (d,p); 1/2 to (s,s), (d,d), and (s,d)

[a/] Symmetric; if -2/3 to (s,p), then -2/3 to (p,s), etc.

Note that **p**, **s**, and **d** will be ordered animal numbers (from one to the last animal) and that each (p,p), (s,p), etc., combination is a location in the inverse of the relationship matrix.

In the example that follows, three animals are base animals: GS1, D1, and GS2. The other five animals are related through those three animals. Animals GS1, D1, and GS2 must be included in building the inverse even if predictions of breeding values are wanted only for S1, . . . , S5.

The relationships for the example are as diagrammed:



The table will be computed beginning with the information for S1 where a blank indicates the parent is not known or does not need to be included as a base animal.

Animal p	Sire s	Dam d	What is added to where (symmetric)
S1	GS1	----	4/3 to (S1,S1);-2/3 to (GS1,S1); 1/3 to (GS1,GS1)
S2	GS1	D1	2 to (S2,S2);-1 to (GS1,S2) and (D1,S2); 1/2 to (GS1,GS1), (D1,D1) and (GS1,D1)
S3	S1	----	4/3 to (S3,S3);-2/3 to (S1,S3);1/3 to (S1,S1)
S4	GS2	D1	2 to (S4,S4);-1 to (GS2,S4) and (D1,S4); 1/2 to (GS2,GS2), (D1,D1) and (GS2,D1)
S5	GS2	D1	2 to (S5,S5);-1 to (GS2,S5) and (D1,S5); 1/2 to (GS2,GS2), (D1,D1) and (GS2,D1)
GS1	----	----	1 to (GS1,GS1)
D1	----	----	1 to (D1,D1)
GS2	----	----	1 to (GS2,GS2)

Note that the base animals are included with unknown parents so that a one is added to the diagonal corresponding to each base animal.

After all eight animals have been processed the inverse of the additive relationship matrix is:

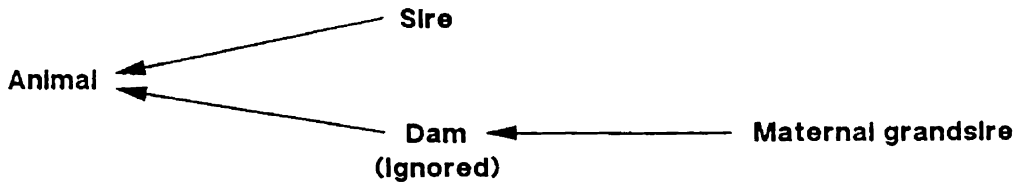
	S1	S2	S3	S4	S5	GS1	D1	GS2
S1	10/6	0	-4/6	0	0	-4/6	0	0
S2		12/6	0	0	0	-6/6	-6/6	0
S3			8/6	0	0	0	0	0
S4				12/6	0	0	-6/6	-6/6
S5					12/6	0	-6/6	-6/6
GS1						11/6	3/6	0
D1			Symmetric				15/6	6/6
GS2								12/6

When animals are inbred the procedure is more complicated although Quaas (1976) developed a method of computing the diagonals of the relationship matrix from which the inverse of the whole additive relationship matrix can be computed relatively easily although not as easily as when inbreeding is ignored. The animals must be ordered by age when inbreeding is considered.

A similar set of rules can be developed for relationships among males; i.e., when the maternal grandsire is used in the calculation rather than the dam. Most relationships in a population are due to males rather than to females.

RULES WHEN SIRES AND MATERNAL GRANDSIRES ARE KNOWN

Most relationships among sires that are evaluated from progeny records are due to male relatives because few dams have more than one son with progeny. Rules for building the inverse of the relationship matrix from known sire and maternal grandsire are similar to those using known sire and dam. Base animals with more than one related collateral descendent must be included as before. With the following rules inbreeding is again ignored.



If known			
Animal p	Sire s	Maternal grandsire m	Then add what to where (symmetric)
Yes	No	No	1 to (p,p)
Yes	Yes	No	4/3 to (p,p); 1/3 to (s,s); -2/3 to (s,p)
Yes	No	Yes	16/15 to (p,p); 1/15 to (m,m); -4/15 to (m,p)
Yes	Yes	Yes	16/11 to (p,p); -8/11 to (s,p); -4/11 to (m,p) 4/11 to (s,s); 2/11 to (m,s); 1/11 to (m,m)

CHAPTER 28

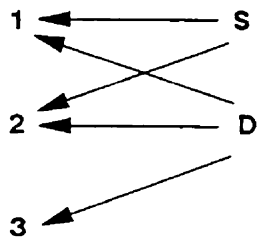
MODELS WITH ANIMALS RELATED

THE ANIMAL MODEL WITH ANIMALS RELATED

The selection index takes advantage of records of relatives to improve predictions. Records of relatives are partial replicates of some of the same genetic effects. The mixed model procedure can also be improved by using the numerator relationships for partial replication. Instead of adding σ_w^2/σ_g^2 to the diagonal of least squares equations of each **g** equation, a function of the additive relationship matrix and $\sigma_w^2/\sigma_g^2 = \lambda$ is added to the block of coefficients for the **g** equations. The additive relationship table can be considered as a matrix of additive relationships with the symbol, A. The function of A used in the mixed model equations is its inverse, A^{-1}, multiplied by the scalar, σ_w^2/σ_g^2.

For this example, the least squares and mixed model equations will be written in matrix notation.

Assume animals 1, 2, 3 each have a record and are related through S and D as diagrammed:



Thus, the additive relationships among animals 1, 2, and 3 are:

$$A = \begin{pmatrix} 1 & 1/2 & 1/4 \\ 1/2 & 1 & 1/4 \\ 1/4 & 1/4 & 1 \end{pmatrix} \quad \text{with } A^{-1} = \begin{pmatrix} 15/11 & -7/11 & -2/11 \\ -7/11 & 15/11 & -2/11 \\ -2/11 & -2/11 & 12/11 \end{pmatrix} .$$

Let y_1, y_2, y_3 be single records of the 3 animals. When the only fixed effect in the model is μ, the least squares equations are:

$$\begin{pmatrix} 3 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{g}_1 \\ \hat{g}_2 \\ \hat{g}_3 \end{pmatrix} = \begin{pmatrix} y_. \\ y_1 \\ y_2 \\ y_3 \end{pmatrix} .$$

To convert the least squares equations to mixed model equations, λA^{-1} is added to the block of coefficients for the g equations. For example, if $\sigma_w^2/\sigma_g^2 = (1-h^2)/h^2 = 3$, then

$$\lambda A^{-1} = 3 \begin{pmatrix} 15/11 & -7/11 & -2/11 \\ -7/11 & 15/11 & -2/11 \\ -2/11 & -2/11 & 12/11 \end{pmatrix}$$

and the mixed model equations become:

$$\begin{pmatrix} 3 & 1 & 1 & 1 \\ 1 & 1+45/11 & 0-21/11 & 0-6/11 \\ 1 & 0-21/11 & 1+45/11 & 0-6/11 \\ 1 & 0-6/11 & 0-6/11 & 1+36/11 \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{g}_1 \\ \hat{g}_2 \\ \hat{g}_3 \end{pmatrix} = \begin{pmatrix} y_. \\ y_1 \\ y_2 \\ y_3 \end{pmatrix} .$$

EQUATIONS AUGMENTED FOR RELATIVES WITHOUT RECORDS

Calculation of A with many animals is difficult. After A has been calculated, the calculation of A^{-1} for many animals is usually prohibitive because computing time for A^{-1} from A is proportional to $\mathbf{n}^3$ where **n** is the number of animals. In 1975 C. R. Henderson made a remarkable discovery that allows rapid and direct calculation of elements of A^{-1} without calculation of A. (See Chapter 27 for rules for calculation of A^{-1}.) The method, however, requires including in A^{-1} the ancestors that create the relationships. In the previous example, S and D as well as animals 1, 2, and 3 must be included in A^{-1}.

To use the rules for rapid calculation of A^{-1}, the mixed model equations for animals with records are augmented with equations for the ancestors without records (also a result due to C. R. Henderson). Let A_+^{-1} be the inverse of A_+ which includes the ancestors without records that create relationships among the animals with records. The right-hand sides of the least squares equations for animals without records are all zero (the model for zero is zero) as are the coefficients of the least squares equations. The **g**'s for the ancestors are included in the solution vector. When λA_+^{-1} is added to the block of coefficients for the animals including the ancestors without records, the coefficients are not all zero for the ancestor equations although the right-hand sides are zero.

The equations for animals with records and the equations for their ancestors without records are tied together by the inverse of the full relationship matrix.

The procedure will be illustrated with the previous example. The least squares equations with equations for g_S and g_D added are:

$$\begin{pmatrix} 3 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{g}_1 \\ \hat{g}_2 \\ \hat{g}_3 \\ \hat{g}_S \\ \hat{g}_D \end{pmatrix} = \begin{pmatrix} y_. \\ y_1 \\ y_2 \\ y_3 \\ 0 \\ 0 \end{pmatrix} .$$

Using the rules for calculating A_+^{-1} with S and D included at the end to agree with the order of the solution vector of the least squares equations:

$$A_+^{-1} = \begin{pmatrix} 2 & 0 & 0 & -1 & -1 \\ 0 & 2 & 0 & -1 & -1 \\ 0 & 0 & 4/3 & 0 & -2/3 \\ -1 & -1 & 0 & 2 & 1 \\ -1 & -1 & -2/3 & 1 & 7/3 \end{pmatrix} .$$

Then λA_+^{-1} is added to the block of coefficients corresponding to the equations for the g's.

A somewhat surprising result is that the solutions for $\hat{\mu}$, $\hat{g}_1$, $\hat{g}_2$, and $\hat{g}_3$ from the augmented equations are exactly the same as when λA^{-1} is added to equations for g_1, g_2, and g_3 in the previous example. That result becomes less unexpected after considering that exactly the same records are available and relationships in both cases were calculated using S and D. Even though more equations must be solved with the augmented procedure, the total computing time is usually much less than calculating A, then A^{-1} for animals with records and finally solving the equations. If, in the augmented equations, the equations for g_S and g_D are absorbed into the equations for g_1, g_2, and g_3, the equations will be identical

to those set up directly for g_1, g_2, and g_3. Thus, mathematically, the solutions must be equal. The solutions for g_S and g_D are based on their relationships to the animals with records as will be illustrated with a simpler example. Assume C has **n** records with parents S and D not having records. With animals ordered C, S, D:

$$A_+^{-1} = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 3/2 & 1/2 \\ -1 & 1/2 & 3/2 \end{pmatrix} .$$

For the model: $y_{ij} = \mu + g_i + p_i + w_{ij}$ with $\lambda = \sigma_w^2/\sigma_g^2 = (1\text{-}r)/h^2$ and $\gamma = \sigma_w^2/\sigma_p^2 = (1\text{-}r)/(r\text{-}h^2)$, the augmented equations are:

$$\begin{pmatrix} n & n & n & 0 & 0 \\ n & n+\gamma & n & 0 & 0 \\ n & n & n+2\lambda & -\lambda & -\lambda \\ 0 & 0 & -\lambda & 3\lambda/2 & \lambda/2 \\ 0 & 0 & -\lambda & \lambda/2 & 3\lambda/2 \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{p}_C \\ \hat{g}_C \\ \hat{g}_S \\ \hat{g}_D \end{pmatrix} = \begin{pmatrix} y_{1.} \\ y_{1.} \\ y_{1.} \\ 0 \\ 0 \end{pmatrix}$$

Note that $\hat{g}_S = \hat{g}_D$ because each has the same relationship to their progeny which has records. Let a parent solution be $\hat{g}_P$. Then from either of the last two equations:

$$(3/2 + 1/2)\lambda\hat{g}_P = \lambda\hat{g}_C$$

so that as might be expected:

$$\hat{g}_P = \hat{g}_C/2 \quad .$$

Substitute $\hat{g}_C/2$ for $\hat{g}_S$ and $\hat{g}_D$ in the equation for g_C and:

$$n\hat{\mu} + n\hat{p}_C + (n+2\lambda)\hat{g}_C - (\lambda/2)\hat{g}_C - (\lambda/2)\hat{g}_C = y_{1.}$$

This equation, on combining terms, is the same as the equation for g_C if relationships to parents S and D had been ignored: $n\hat{\mu} + n\hat{p}_C + (n+\lambda)\hat{g}_C = y_{1.}$

This result is expected because S and D did not contribute information to evaluate C.

In general, the reasons to include animals with no records are 1) that such so-called base animals establish relationships among animals with records, e.g., if a full sib of C had a record, S and D create the relationships needed to use that record in the evaluation for C and 2) that calculating A_+^{-1} is much, much easier than calculating A and then A^{-1} for animals with records.

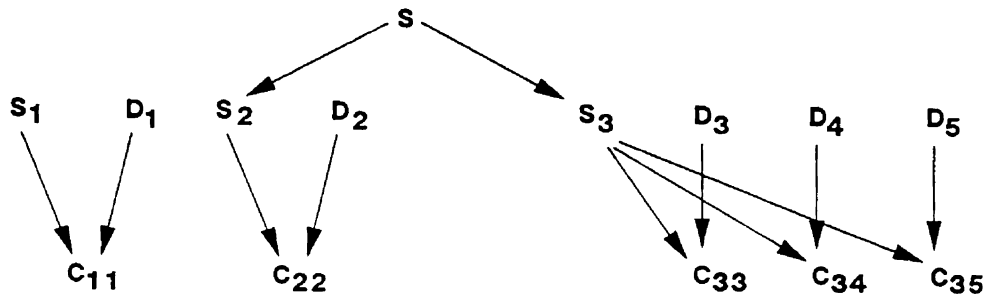
This example suggests that an animal model can be used to evaluate sires that have many progeny. In the simple example the evaluation of S was obtained and was based on the records of only one progeny.

SIRE EVALUATION WITH ANIMAL MODEL

The animal model is ideal for evaluating sires from their progeny records because the merit of the mates of the sires (dams of the progeny) will be adjusted for automatically through the relationship matrix in the same way that the selection index can be used to account for association of some sires with better than average mates and other sires with poorer than average mates. The disadvantage of the animal model for large data sets is that a large number of equations must be solved. The number of equations is somewhat larger than the number of animals and depends on whether repeated records are used, how many animals without records are included (base and sires) and on the number of other factors in the model to account for such factors as management and seasonal effects.

In the augmented procedure, a base animal that has only one relative to be evaluated does not have to be included in A_+^{-1}, e.g, a sire that has a son with progeny but has no progeny with records nor any other sons with progeny or descendants in the group of animals to be evaluated can be, but does not need to be, included in the augmented equations.

The example which follows shows the equations for sire evaluation with an animal model for the situation where all of the mates of sires are unrelated to each other and to the sires. Thus, parents of the mates are assumed not to have records and do not need to be included in A_+^{-1}. They could be included and the solutions would be the same for the other animals but more equations would need to be solved. To further simplify the example, only females will have records and each will have a single record.



All C and D animals have a single record.

There will be:

10 equations for animals with records,

4 equations for sires including S that has two sons with progeny, and

1 equation for μ for the simple animal model.

Let the animals be ordered:

$D_1, D_2, D_3, D_4, D_5, C_{11}, C_{22}, C_{33}, C_{34}, C_{35}, S, S_1, S_2, S_3$

The mixed model equations for the simple animal model with $\sigma_w^2/\sigma_g^2 = (1\text{-}h^2)/h^2 = \lambda$ are shown on the next page.

$$
\begin{pmatrix}
10 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
 & 1+3\lambda/2 & 0 & 0 & 0 & 0 & -\lambda & 0 & 0 & 0 & 0 & 0 & \lambda/2 & 0 & 0 \\
 & & 1+3\lambda/2 & 0 & 0 & 0 & 0 & -\lambda & 0 & 0 & 0 & 0 & 0 & \lambda/2 & 0 \\
 & & & 1+3\lambda/2 & 0 & 0 & 0 & 0 & -\lambda & 0 & 0 & 0 & 0 & 0 & \lambda/2 \\
 & & & & 1+3\lambda/2 & 0 & 0 & 0 & 0 & -\lambda & 0 & 0 & 0 & 0 & \lambda/2 \\
 & & & & & 1+3\lambda/2 & 0 & 0 & 0 & 0 & -\lambda & 0 & 0 & 0 & \lambda/2 \\
 & & & & & & 1+2\lambda & 0 & 0 & 0 & 0 & 0 & -\lambda & 0 & 0 \\
 & & & & & & & 1+2\lambda & 0 & 0 & 0 & 0 & 0 & -\lambda & 0 \\
 & & & & & & & & 1+2\lambda & 0 & 0 & 0 & 0 & 0 & -\lambda \\
 & & & & & & & & & 1+2\lambda & 0 & 0 & 0 & 0 & -\lambda \\
 & & & & & & & & & & 1+2\lambda & 0 & 0 & 0 & -\lambda \\
 & & & & & & & & & & & 5\lambda/3 & 0 & -2\lambda/3 & -2\lambda/3 \\
 & & & & & & & & & & & & 3\lambda/2 & 0 & 0 \\
 & & & & & & & & & & & & & 11\lambda/6 & 0 \\
 & & & & & & & & & & & & & & 17\lambda/6
\end{pmatrix}
\begin{pmatrix} \hat{\mu} \\ \hat{g}_{D_1} \\ \hat{g}_{D_2} \\ \hat{g}_{D_3} \\ \hat{g}_{D_4} \\ \hat{g}_{D_5} \\ \hat{g}_{C_{11}} \\ \hat{g}_{C_{22}} \\ \hat{g}_{C_{33}} \\ \hat{g}_{C_{34}} \\ \hat{g}_{C_{35}} \\ \hat{g}_{S} \\ \hat{g}_{S_1} \\ \hat{g}_{S_2} \\ \hat{g}_{S_3} \end{pmatrix}
=
\begin{pmatrix} y_{\cdot} \\ y_{D_1} \\ y_{D_2} \\ y_{D_3} \\ y_{D_4} \\ y_{D_5} \\ y_{C_{11}} \\ y_{C_{22}} \\ y_{C_{33}} \\ y_{C_{34}} \\ y_{C_{35}} \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}
$$

SYMMETRIC
(coefficients in a column below the diagonal are the same as those in the row to the right of the diagonal)

Examination of the solution for a sire shows the weight for each mate is minus one-half (-1/2) of that for each progeny. In other words, an estimate of one-half of the dam's genetic value is subtracted from the estimated genetic value of the progeny to account for the mate's merit to leave only the part of the progeny's genetic value contributed by the sire. For example, for S_3:

$$(17\lambda/6)\hat{g}_{S_3} = (2\lambda/3)\hat{g}_S + \lambda(\hat{g}_{C_{33}} + \hat{g}_{C_{34}} + \hat{g}_{C_{35}}) - (\lambda/2)(\hat{g}_{D_3} + \hat{g}_{D_4} + \hat{g}_{D_5})$$

$$(17\lambda/6)\hat{g}_{S_3} = (2\lambda/3)\hat{g}_S + \lambda[(\hat{g}_{C_{33}} - \hat{g}_{D_3}/2) + (\hat{g}_{C_{34}} - \hat{g}_{D_4}/2) + (\hat{g}_{C_{35}} - \hat{g}_{D_5}/2)]$$

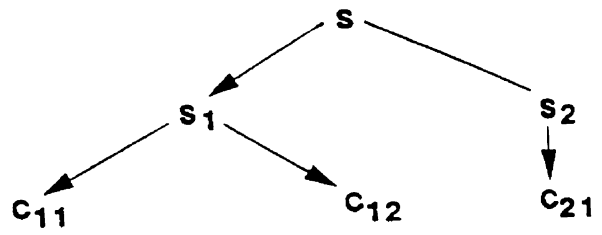
And for S_1:

$$\hat{g}_{S_1} = (2/3)(\hat{g}_{C_{11}} - \hat{g}_{D_1}/2) \ .$$

SIRE EVALUATION WITH ANIMAL MODEL IGNORING MATES AND RELATIONSHIPS THROUGH FEMALES

In the past, sire evaluations generally were done ignoring records on mates because of the computing time required. The animal model can be used with the same approximation; that all mates are unrelated to each other and to the sires. This approximation to the full animal model is equivalent to assuming that only relationships from males to males are important.

Consider the following example where C_{11}, C_{12}, and C_{21} have single records, y_{11}, y_{12}, and y_{21}.



In calculating A_+^{-1}, all dams are considered to be unknown so that for animals ordered C_{11}, C_{12}, C_{21}, S_1, S_2, and S (notice that each term in A_+^{-1} when the dam is missing contains a 3 in the denominator):

$$A_+^{-1} = (1/3)\begin{pmatrix} 4 & 0 & 0 & -2 & 0 & 0 \\ 0 & 4 & 0 & -2 & 0 & 0 \\ 0 & 0 & 4 & 0 & -2 & 0 \\ -2 & -2 & 0 & 6 & 0 & -2 \\ 0 & 0 & -2 & 0 & 5 & -2 \\ 0 & 0 & 0 & -2 & -2 & 5 \end{pmatrix}.$$

For $\sigma_w^2/\sigma_g^2 = \lambda$, the mixed model equations are:

$$\begin{pmatrix} 3 & 1 & 1 & 1 & 0 & 0 & 0 \\ & 1+4\lambda/3 & 0 & 0 & -2\lambda/3 & 0 & 0 \\ & & 1+4\lambda/3 & 0 & -2\lambda/3 & 0 & 0 \\ & & & 1+4\lambda/3 & 0 & -2\lambda/3 & 0 \\ & & & & 6\lambda/3 & 0 & -2\lambda/3 \\ & & & & & 5\lambda/3 & -2\lambda/3 \\ \text{Symmetric} & & & & & & 5\lambda/3 \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{g}_{11} \\ \hat{g}_{12} \\ \hat{g}_{21} \\ \hat{g}_{S_1} \\ \hat{g}_{S_2} \\ \hat{g}_S \end{pmatrix} = \begin{pmatrix} y_{..} \\ y_{11} \\ y_{12} \\ y_{21} \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

The solutions predict genetic values simultaneously for animals with records (the progeny) ignoring relationships arising from females, and for animals without records, in this case, the sires.

CHAPTER 29

SIRE MODELS WITH SOME RELATIONSHIPS

IGNORING MATES AND FEMALE RELATIONSHIPS

The approximate animal model described in the last section of Chapter 28 that ignores relationships through females requires an equation for each progeny. An equation is included for each animal which in some cases may be many. The number of equations can be reduced essentially to the number of sires by using the sire model. With the sire model, only male to male relationships will be considered (assumes dams are unrelated to sires and to each other). The sire model is the same as in Chapter 26 except that now relationships among sires are used:

$$y_{ij} = \mu + s_i + w_{ij}$$

where $s_i = g_i/2$ is the transmitting ability of sire i. Note that $\sigma_s^2 = \sigma_g^2/4$ and $\sigma_w^2 = \sigma_y^2 - \sigma_g^2/4$ or equivalently $\sigma_s^2 = h^2\sigma_y^2/4$ and $\sigma_w^2 = (1 - h^2/4)\sigma_y^2$ so that $\sigma_w^2/\sigma_s^2 = (4 - h^2)/h^2 = \gamma$. The previous example from Chapter 28 will be used. With the sire model, only relationships among S_1, S_2, and S will be considered in calculating A_+^{-1}:

$$A_+^{-1} = (1/3)\begin{pmatrix} 4 & 0 & -2 \\ 0 & 4 & -2 \\ -2 & -2 & 5 \end{pmatrix} .$$

With the sire model, γA_+^{-1} is added to the block of the coefficient matrix corresponding to the sire transmitting abilities. The least squares equations are augmented by equations for sires that have no progeny with records but which create relationships among sires with progeny with records. The mixed model equations for the example are:

$$\begin{pmatrix} 3 & 2 & 1 & 0 \\ & 2+4\gamma/3 & 0 & -2\gamma/3 \\ & & 1+4\gamma/3 & -2\gamma/3 \\ \text{Symmetric} & & & 5\gamma/3 \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_S \end{pmatrix} = \begin{pmatrix} y_{..} \\ y_{1.} \\ y_{2.} \\ 0 \end{pmatrix} .$$

The solutions for $\hat{s}_1$, $\hat{s}_2$, and $\hat{s}_S$ are exactly one-half those for the previous example ($\hat{g}_{S_1}$, $\hat{g}_{S_2}$, and $\hat{g}_S$) that had equations for each progeny and augmented equations for the sires that had no records themselves. That $\hat{s}_1 = \hat{g}_{S_1}/2$ can be shown by absorbing equations for g_{11} and g_{12} in the last section of Chapter 28 into the equation for g_{S_1}. The equation for g_{S_1} will be the same as for s_1 except that the coefficients for $\hat{g}_{S_1}$ and $\hat{g}_S$ are one-half as large as the coefficients for $\hat{s}_1$ and $\hat{s}_S$. The advantage of the sire model as compared to the equivalent approximate animal model is that many fewer equations need to be set up and solved.

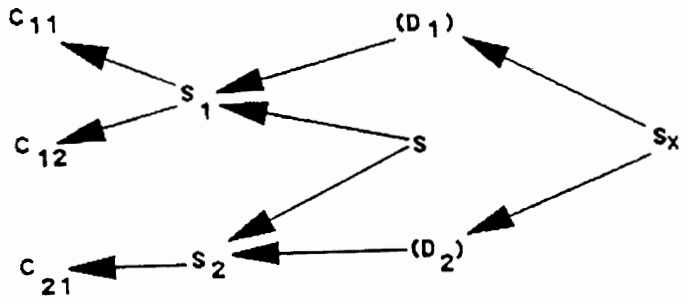
RELATIONSHIPS FROM SIRES AND MATERNAL GRANDSIRES OF MALES

Most relationships among males arise from male ancestors. Even if dams of males are not included in calculation of A_+^{-1} among males, sires of the dams (maternal grandsires of males) can be used in calculation of A_+^{-1} and if they have no progeny with records can be evaluated from the augmented equations. The increased ties among males will result in

slightly increased accuracies of evaluation. Only maternal grandsires that have more than one male descendent in the list to be evaluated or those with progeny with records need to be included in calculating A_+^{-1}. Rules developed by C. R. Henderson for calculating A_+^{-1} from sires and maternal grandsires are similar to the rules using sires and dams, and are given in Chapter 27 for calculating A_+^{-1}.

For this sire model, $\sigma_w^2/\sigma_s^2 = \gamma$.

Assume as an example the same animals and records as in the previous example except that S_x is the maternal grandsire of both S_1 and S_2:



Females D_1 and D_2 are not included in A_+^{-1}. With males ordered S_1, S_2, S, S_x:

$$A_+^{-1} = (1/11)\begin{pmatrix} 16 & 0 & -8 & -4 \\ 0 & 16 & -8 & -4 \\ -8 & -8 & 19 & 4 \\ -4 & -4 & 4 & 13 \end{pmatrix}.$$

For the sire model and with $\sigma_w^2/\sigma_s^2 = (4 - h^2)/h^2 = \gamma$; the elements of γA_+^{-1} are added to the coefficients of the least squares equations corresponding to equations for S_1, S_2 and the augmented equations for S and S_x. The augmented mixed model equations are:

$$\begin{pmatrix} 3 & 2 & 1 & 0 & 0 \\ & 2+16\gamma/11 & 0 & -8\gamma/11 & -4\gamma/11 \\ & & 1+16\gamma/11 & -8\gamma/11 & -4\gamma/11 \\ & & & 19\gamma/11 & 4\gamma/11 \\ \text{Symmetric} & & & & 13\gamma/11 \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_S \\ \hat{s}_{S_x} \end{pmatrix} = \begin{pmatrix} y_{..} \\ y_{1.} \\ y_{2.} \\ 0 \\ 0 \end{pmatrix}$$

The variance of prediction errors, e.g., $V(\hat{s}_i - s_i)$, will be decreased as compared to ignoring S_x.

CHAPTER 30

VARIANCE OF PREDICTION ERRORS

Calculation of $r_{T\hat{T}}$ and variance of prediction error, $V(T - \hat{T}) = (1 - r^2_{T\hat{T}})\sigma^2_T$ can be done for solutions from mixed model equations as from selection index theory. How to do these calculations will be described for models including: genetic value, i.e., when T = g; transmitting ability, when T = s; and producing ability, when T = g + p; as well as for variances of estimates of fixed effects estimated from data. The calculation for $V(T - \hat{T})$ first requires the inverse of the coefficient matrix for the mixed model equations and σ^2_w. The second step is to calculate $r^2_{T\hat{T}}$ from $V(T - \hat{T})$ which will also require the ratio of σ^2_w/σ^2_T. For example, assume the repeated records model:

$$y_{ij} = \mu + g_i + p_i + w_{ij} .$$

Let $\lambda = \sigma^2_w/\sigma^2_g$ and $\gamma = \sigma^2_w/\sigma^2_p$. The symbolic mixed model equations are:

$$\begin{pmatrix} c_{\mu,\mu} & c_{\mu,p_1} & \cdots & c_{\mu,g_1} & \cdots \\ c_{p_1,\mu} & c_{p_1,p_1} & \cdots & c_{p_1,g_1} & \cdots \\ \vdots & & & \vdots & \\ c_{g_1,\mu} & c_{g_1,p_1} & \cdots & c_{g_1,g_1} & \cdots \\ \vdots & & & \vdots & \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{p}_1 \\ \vdots \\ \hat{g}_1 \\ \vdots \end{pmatrix} = \begin{pmatrix} y_{..} \\ y_{1.} \\ \vdots \\ y_{1.} \\ \vdots \end{pmatrix}$$

where $c_{\mu,\mu} = n_.$, $c_{\mu,p_1} = n_1$, $c_{\mu,g_1} = n_1$, $c_{p_1,p_1} = n_1 + \gamma$

and the c_{gg} block = $\begin{pmatrix} n_1 & 0 & \cdots \\ 0 & n_2 & \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \end{pmatrix} + \lambda A^{-1}$

In matrix terms $\mathbf{Cs} = \mathbf{r}$ where the solution vector: $\mathbf{s} = \mathbf{C}^{-1}\mathbf{r}$.

Let the elements of the inverse of **C** be:

$$\mathbf{C}^{-1} = \begin{pmatrix} c^{\mu,\mu} & c^{\mu,p1} & \cdots & c^{\mu,g1} & \cdots \\ c^{p1,\mu} & c^{p1,p1} & \cdots & c^{p1,g1} & \cdots \\ \cdot & & & & \\ \cdot & & & & \\ c^{g1,\mu} & c^{g1,p1} & \cdots & c^{g1,g1} & \cdots \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \end{pmatrix} .$$

These terms when multiplied by σ_w^2 correspond to prediction error variances and covariances:

$$V(g_1 - \hat{g}_1) = c^{g1,g1}\sigma_w^2, \quad V(p_1 - \hat{p}_1) = c^{p1,p1}\sigma_w^2 \text{ and}$$

$$\text{Cov}(g_1 - \hat{g}_1, p_1 - \hat{p}_1) = c^{g1,p1}\sigma_w^2$$

so that for producing ability;

$$V[(g_1 + p_1) - (\hat{g}_1 + \hat{p}_1)] = V(g_1 - \hat{g}_1) + V(p_1 - \hat{p}_1) + 2\text{Cov}(g_1 - \hat{g}_1, p_1 - \hat{p}_1)$$

$$= (c^{g1,g1} + c^{p1,p1} + 2c^{g1,p1})\sigma_w^2 .$$

VARIANCE OF PREDICTION ERROR OF GENETIC VALUE

Because $V(g_i - \hat{g}_i) = c^{gi,gi}\sigma_w^2$ and also from the selection index property $V(g_i - \hat{g}_i) = (1 - r_{g\hat{g}}^2)\sigma_g^2$, then $c^{gi,gi}\sigma_w^2 = (1 - r_{g\hat{g}}^2)\sigma_g^2$.

The equation can be solved for $r_{g\hat{g}}^2$ as: $r_{g\hat{g}}^2 = 1 - c^{gi,gi}(\sigma_w^2/\sigma_g^2)$.

The ratio, $\sigma_w^2/\sigma_g^2 = (1 - r)/h^2$, is known and is used in calculating **C**. The other term, $c^{gi,gi}$, is the appropriate element from the inverse of C. If animal **i** is inbred, then the ratio σ_w^2/σ_g^2 becomes $\sigma_w^2/\sigma_g^2 (1 + F_i)$.

VARIANCE OF PREDICTION ERROR OF PRODUCING ABILITY

Because from mixed model theory

$$V[(g_i + p_i) - (\hat{g}_i + \hat{p}_i)] = (c^{gi,gi} + c^{pi,pi} + 2c^{gi,pi})\sigma_w^2$$ equals

$(1 - r_{g+p,\hat{g}+\hat{p}}^2)(\sigma_g^2 + \sigma_p^2)$ from selection index theory, then:

$$r_{g+p,\hat{g}+\hat{p}}^2 = 1 - (c^{gi,gi} + c^{pi,pi} + 2c^{gi,pi})\sigma_w^2/(\sigma_g^2 + \sigma_p^2).$$

The ratio $\sigma_w^2/(\sigma_g^2 + \sigma_p^2) = (1 - r)/r$ is known when calculating **C**. The other terms are appropriate elements of the inverse of **C**.

VARIANCE OF PREDICTION ERROR FOR OTHER MODELS

For other models, calculations are similar to those described in previous paragraphs. The appropriate elements from $\mathbf{C}^{-1}$ are multiplied by σ_w^2 for that model. For the sire model, $\sigma_w^2 = (1 - h^2/4)\sigma_y^2$ and $\sigma_w^2/\sigma_s^2 = (4 - h^2)/h^2$. For the animal model with a single record per animal, $\sigma_w^2 = (1 - h^2)\sigma_y^2$ and $\sigma_w^2/\sigma_g^2 = (1 - h^2)/h^2$.

VARIANCE OF PREDICTION ERROR FOR MODELS WITH MORE FIXED EFFECTS

The elements of $\mathbf{C}^{-1}$ corresponding to random effects such as, **g**, **p**, and **s**, are used for calculating variances and covariances of prediction errors for those effects.

These inverse elements, however, depend partially on the fixed effects in the model and on the distribution of records among the levels of the fixed factors. With several fixed factors in the model, **C** is singular so that an inverse cannot be obtained. If constraints are imposed so that the constrained $\mathbf{C}_*$ is nonsingular, then $\mathbf{C}_*^{-1}$ can be obtained but will depend on the set of constraints chosen. The expected values of solutions for fixed effects depend on the constraints.

Nevertheless, solutions for random effects, such as **g**, **p**, and **s**, will be the same for any set of permissible constraints. Similarly, prediction error variances for the random effects do not depend on the constraints chosen, i.e., the block of elements of $\mathbf{C}_*^{-1}$ corresponding to the random effects is unique and does not depend on the constraints.

VARIANCES OF ESTIMATES OF FIXED EFFECTS

Estimates of fixed effects also have variances. For example, the variance of $\hat{\mu}$ is $c^{\mu,\mu}\sigma_w^2$ for models in which μ is the only fixed effect. For models with more fixed factors, the variances of the estimates are determined similarly from the inverse of $\mathbf{C}_*^{-1}$. The problem, however, is that because of the constraints needed to obtain solutions, the expected values of the solutions are not the effects represented by $\hat{\mu}$, etc. Generally differences between levels of a factor are estimable. For example, depending on the model and constraints, $E[\hat{f}_1 - \hat{f}_2]$ may equal $f_1 - f_2$ for levels 1 and 2 of fixed factor f. Then the variance of the estimable difference, $\hat{f}_1 - \hat{f}_2$ is $V(\hat{f}_1 - \hat{f}_2) = (c^{f1,f1} + c^{f2,f2} - 2c^{f1,f2})\sigma_w^2$.

CHAPTER 31

NUMERICAL EXAMPLE OF ANIMAL MODEL WITH DIFFERENT CONSTRAINTS

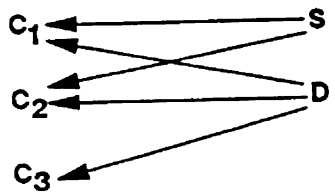
This chapter will demonstrate the effect of constraints on solutions to the mixed model equations for a repeated records animal model.

The model for a record k of animal j affected by level i of fixed factor f is:

$$y_{ijk} = \mu + f_i + p_j + g_j + w_{ijk} \ .$$

Let $\sigma^2_w = (2000)^2$ and with r = .6 and $h^2 = .4$, $\sigma^2_w/\sigma^2_g = (1 - r)/h^2 = \lambda = 1$ and $\sigma^2_w/\sigma^2_p = (1 - r)/(r - h^2) = \gamma = 2$.

The animals with records are C_1 (2 records), C_2 (1 record), and C_3 (3 records). The parents of C_1 and C_2 are S and D, and one parent of C_3 is D with the other parent unknown and not needed because it has only one relative with a record. (See Chapter 28 for example of animal model with equations augmented for relatives without records.)



AUGMENTED MIXED MODEL EQUATIONS

The records for the animals are distributed in the levels of the fixed factor as follows:

Fixed factor	Animal 1	2	3	Fixed factor totals
f_1	10,000	----	9,000	19,000
f_2	12,000	----	10,000	22,000
f_3	----	15,000	12,000	27,000
Totals	22,000	15,000	31,000	68,000

The augmented mixed model equations are:

$$\begin{pmatrix}
6 & 2 & 2 & 2 & 2 & 1 & 3 & 2 & 1 & 3 & 0 & 0 \\
 & 2 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\
 & & 2 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\
 & & & 2 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\
 & & & & 2+2 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\
 & & & & & 1+2 & 0 & 0 & 1 & 0 & 0 & 0 \\
 & & & & & & 3+2 & 0 & 0 & 3 & 0 & 0 \\
 & & & & & & & 2+2 & 0 & 0 & -1 & -1 \\
 & & & & & & & & 1+2 & 0 & -1 & -1 \\
 & & \text{symmetric} & & & & & & & 3+(4/3) & 0 & -2/3 \\
 & & & & & & & & & & +2 & +1 \\
 & & & & & & & & & & & +2(1/3)
\end{pmatrix}
\begin{pmatrix} \hat{\mu} \\ \hat{f}_1 \\ \hat{f}_2 \\ \hat{f}_3 \\ \hat{p}_1 \\ \hat{p}_2 \\ \hat{p}_3 \\ \hat{g}_1 \\ \hat{g}_2 \\ \hat{g}_3 \\ \hat{g}_S \\ \hat{g}_D \end{pmatrix}
=
\begin{pmatrix} 68{,}000 \\ 19{,}000 \\ 22{,}000 \\ 27{,}000 \\ 22{,}000 \\ 15{,}000 \\ 31{,}000 \\ 22{,}000 \\ 15{,}000 \\ 31{,}000 \\ 0 \\ 0 \end{pmatrix}$$

CONSTRAINTS

One constraint will be needed because the f equations sum to the μ equation.

With the constraint $\hat{f}_1 = 0$, the inverse of the coefficient matrix is obtained by zeroing the row and column coefficients for $\hat{f}_1$ and then inverting the remaining matrix.

In the inverse the row and column of zeros for $\hat{f}_1$ are shown:

1.310	.000	-.500	-.657	-.222	-.065	-.213	-.616	-.458	-.569	-.287	-.500
.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
-.500	.000	1.000	.500	.000	.000	.000	.000	.000	.000	.000	.000
-.657	.000	.500	1.296	.139	-.157	.019	.130	-.167	.028	-.019	.000
-.222	.000	.000	.139	.417	.028	.056	-.111	.000	.083	-.056	.000
-.065	.000	.000	-.157	.028	.435	.037	.009	-.083	.056	-.037	.000
-.213	.000	.000	.019	.056	.037	.407	.102	.083	-.139	.093	.000
-.616	.000	.000	.130	-.111	.009	.102	.838	.458	.403	.398	.500
-.458	.000	.000	-.167	.000	-.083	.083	.458	.875	.375	.417	.500
-.569	.000	.000	.028	.083	.056	-.139	.403	.375	.792	.139	.500
-.287	.000	.000	-.019	-.056	-.037	.093	.398	.417	.139	.907	.000
-.500	.000	.000	.000	.000	.000	.000	.500	.500	.500	.000	1.000

With the constraint $\hat{\mu} = 0$, coefficients for the μ row and column are zeroed.

The inverse is:

.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
.000	1.310	.810	.653	-.222	-.065	-.213	-.616	-.458	-.569	-.287	-.500
.000	.810	1.310	.653	-.222	-.065	-.213	-.616	-.458	-.569	-.287	-.500
.000	.653	.653	1.292	-.083	-.222	-.194	-.486	-.625	-.542	-.306	-.500
.000	-.222	-.222	-.083	.417	.028	.056	-.111	.000	.083	-.056	.000
.000	-.065	-.065	-.222	.028	.435	.037	.009	-.083	.056	-.037	.000
.000	-.213	-.213	-.194	.056	.037	.407	.102	.083	-.139	.093	.000
.000	-.616	-.616	-.486	-.111	.009	.102	.838	.458	.403	.398	.500
.000	-.458	-.458	-.625	.000	-.083	.083	.458	.875	.375	.417	.500
.000	-.569	-.569	-.542	.083	.056	-.139	.403	.375	.792	.139	.500
.000	-.287	-.287	-.306	-.056	-.037	.093	.398	.417	.139	.907	.000
.000	-.500	-.500	-.500	.000	.000	.000	.500	.500	.500	.000	1.000

With the constraint $\hat{f}_3 = 0$ the inverse is:

1.292	-.639	-.639	.000	-.083	-.222	-.194	-.486	-.625	-.542	-.306	-.500
-.639	1.296	.796	.000	-.139	.157	-.019	-.130	.167	-.028	.019	.000
-.639	.796	1.296	.000	-.139	.157	-.019	-.130	.167	-.028	.019	.000
.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
-.083	-.139	-.139	.000	.417	.028	.056	-.111	.000	.083	-.056	.000
-.222	.157	.157	.000	.028	.435	.037	.009	-.083	.056	-.037	.000
-.194	-.019	-.019	.000	.056	.037	.407	.102	.083	-.139	.093	.000
-.486	-.130	-.130	.000	-.111	.009	.102	.838	.458	.403	.398	.500
-.625	.167	.167	.000	.000	-.083	.083	.458	.875	.375	.417	.500
-.542	-.028	-.028	.000	.083	.056	-.139	.403	.375	.792	.139	.500
-.306	.019	.019	.000	-.056	-.037	.093	.398	.417	.139	.907	.000
-.500	.000	.000	.000	.000	.000	.000	.500	.500	.500	.000	1.000

PREDICTION ERROR VARIANCES

Notice that with any of the three constraints that the blocks of the inverses corresponding to p_1, p_2, p_3, g_1, g_2, g_3, g_S and g_D are the same. The solutions for those effects are also the same as is shown in Table 31.1. In technical jargon, this means that the predictors of the random effects are invariant to (do not depend on) the choice of constraints.

For example with all three sets of constraints:

$$V(g_1 - \hat{g}_1) = c^{g_1,g_1}\sigma_w^2 = .838(2000)^2.$$

$$V(g_D - \hat{g}_D) = c^{g_D,g_D}\sigma_w^2 = 1.000(2000)^2$$

$$V[(g_1 + p_1) - (\hat{g}_1 + \hat{p}_1)] = (c^{p_1,p_1} + c^{g_1,g_1} + 2c^{p_1,g_1})\sigma_w^2$$

$$= [.417 + .838 + 2(-.111)](2000)^2 \quad .$$

VARIANCES OF ESTIMATES OF FIXED EFFECTS

Because what $\hat{f}_i$ estimates depends on the constraint, $V(\hat{f}_i)$ is different from constraint to constraint although the variance of an estimable function of the **f**'s is the same with any of the constraints.

With all 3 sets of constraints $\hat{f}_2 - \hat{f}_3$ estimates $f_2 - f_3$, i.e., $E[\hat{f}_2 - \hat{f}_3] = f_2 - f_3$. For all three cases, $V(\hat{f}_2 - \hat{f}_3)$ are the same.

For $\hat{f}_1 = 0$:

$$V(\hat{f}_2 - \hat{f}_3) = (c^{f_2,f_2} + c^{f_3,f_3} - 2c^{f_2,f_3})\sigma_w^2$$

$$= [1 + 1.296 - 2(.500)]\sigma_w^2 = 1.296\sigma_w^2$$

For $\hat{\mu} = 0$:

$$V(\hat{f}_2 - \hat{f}_3) = [1.310 + 1.292 - 2(.653)]\sigma_w^2 = 1.296\sigma_w^2$$

For $\hat{f}_3 = 0$:

$$V(\hat{f}_2 - \hat{f}_3) = [1.296 + 0 - 2(0)]\sigma_w^2 = 1.296\sigma_w^2$$

In the last case, $\hat{f}_3 = 0$. Note that a constant (implied by the constraint) has no variance and similarly the covariance of a constant, $\hat{f}_3 = 0$, with an estimate, $\hat{f}_2$, also is zero.

In all three cases the variance of the estimated difference between f_2 and f_3 is the same, $1.296\sigma_w^2$. From Table 31.1 the estimate of the difference, $f_2 - f_3$, is -2278, i.e., f_3 is estimated to be larger than f_2 by 2278 no matter which constraint is used to obtain a set of solutions.

Table 31.1. Solutions for augmented and nonaugmented mixed model equations with different constraints

Solution	Augmented $\hat{f}_1=0$	Augmented $\hat{\mu}=0$	Augmented $\hat{f}_3=0$	Nonaugmented $\hat{f}_3=0$
$\hat{\mu}$	9,806	0	13,583	13,583
$\hat{f}_1$	0	9,806	-3,778	-3,778
$\hat{f}_2$	1,500	11,306	-2,278	-2,278
$\hat{f}_3$	3,778	13,583	0	0
$\hat{p}_1$	83	83	83	83
$\hat{p}_2$	306	306	306	306
$\hat{p}_3$	-389	-389	-389	-389
$\hat{g}_1$	278	278	278	278
$\hat{g}_2$	500	500	500	500
$\hat{g}_3$	-583	-583	-583	-583
$\hat{g}_S$	389	389	389	----
$\hat{g}_D$	0	0	0	----

EQUIVALENT MIXED MODEL EQUATIONS

If the mixed model equations had not been augmented but S and D had been used to calculate the relationship matrix, A, for C_1, C_2, and C_3, then

$$A = \begin{pmatrix} 1 & 1/2 & 1/4 \\ 1/2 & 1 & 1/4 \\ 1/4 & 1/4 & 1 \end{pmatrix} \text{ and } A^{-1} = \begin{pmatrix} 1.364 & -.636 & -.182 \\ -.636 & 1.364 & -.182 \\ -.182 & -.182 & 1.091 \end{pmatrix} .$$

With $\lambda = 1$, the equivalent mixed model equations are:

$$\begin{pmatrix}
6 & 2 & 2 & 2 & 2 & 1 & 3 & 2 & 1 & 3 \\
 & 2 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\
 & & 2 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\
 & & & 2 & 0 & 1 & 1 & 0 & 1 & 1 \\
 & & & & 4 & 0 & 0 & 2 & 0 & 0 \\
 & & & & & 3 & 0 & 0 & 1 & 0 \\
 & & & & & & 5 & 0 & 0 & 3 \\
 & & & & & & & 3.364 & -.636 & -.182 \\
 & & & & & & & & 2.364 & -.182 \\
 & & & & & & & & & 4.091
\end{pmatrix}
\begin{pmatrix} \hat{\mu} \\ \hat{f}_1 \\ \hat{f}_2 \\ \hat{f}_3 \\ \hat{p}_1 \\ \hat{p}_2 \\ \hat{p}_3 \\ \hat{g}_1 \\ \hat{g}_2 \\ \hat{g}_3 \end{pmatrix}
= \begin{pmatrix} 68{,}000 \\ 19{,}000 \\ 22{,}000 \\ 27{,}000 \\ 22{,}000 \\ 15{,}000 \\ 31{,}000 \\ 22{,}000 \\ 15{,}000 \\ 31{,}000 \end{pmatrix}.$$

The solutions as shown in Table 31.1 are identical to those from the augmented equations. Similarly the variances of prediction errors are also the same as can be seen from the inverse with $\hat{f}_3 = 0$:

$$\begin{pmatrix}
1.292 & -.639 & -.639 & .000 & -.083 & -.222 & -.194 & -.486 & -.625 & -.542 \\
-.639 & 1.296 & .796 & .000 & -.139 & .157 & -.019 & -.130 & .167 & -.028 \\
-.639 & .796 & 1.296 & .000 & -.139 & .157 & -.019 & -.130 & .167 & -.028 \\
.000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 \\
-.083 & -.139 & -.139 & .000 & .417 & .028 & .056 & -.111 & .000 & .083 \\
-.222 & .157 & .157 & .000 & .028 & .435 & .037 & .009 & -.083 & .056 \\
-.194 & -.019 & -.019 & .000 & .056 & .037 & .407 & .102 & .083 & -.139 \\
-.486 & -.130 & -.130 & .000 & -.111 & .009 & .102 & .838 & .458 & .403 \\
-.625 & .167 & .167 & .000 & .000 & -.083 & .083 & .458 & .875 & .375 \\
-.542 & -.028 & -.028 & .000 & .083 & .056 & -.139 & .403 & .375 & .792
\end{pmatrix}.$$

CHAPTER 32

CREATING AND SOLVING LEAST SQUARES AND MIXED MODEL EQUATIONS

ALGORITHM TO CREATE THE LEAST SQUARES EQUATIONS

Computing strategies to accumulate the coefficients and right-hand sides of the least squares equations depend on the amount of data, the model, and computer memory. Nevertheless, a symbolic algorithm can be used to remember which coefficients are involved for each record. Data can be presented for computing one record at a time. Coefficients and right-hand sides associated with each record are summed into computer memory that is assigned and initialized to zero before the first record is processed.

As an example, the model,

$$y_{ijk} = \mu + f_i + p_j + g_j + w_{ijk},$$

has four terms other than the residual. Thus each record is included in four sums corresponding to μ, f_i, p_j and g_j. Each record carries four elements of the model (excluding w_{ijk}) to each sum. Therefore, each record contributes to 16 elements of the coefficient matrix of the left-hand-sides of the least squares equations. The locations in the coefficient matrix, C, can be determined by squaring the model (excluding w_{ijk}):

$$(\mu + f_i + p_j + g_j)^2 .$$

The 16 terms correspond to the subscripts of 16 elements in C where a 1 will be added for that record's contribution to the four sums:

$$\begin{array}{cccc} \mu\mu & \mu f_i & \mu p_j & \mu g_j \\ f_i\mu & f_i f_i & f_i p_j & f_i g_j \\ p_j\mu & p_j f_i & p_j p_j & p_j g_j \\ g_j\mu & g_j f_i & g_j p_j & g_j g_j \end{array} .$$

For example, the diagonals $c_{\mu\mu}$, $c_{f_i f_i}$, $c_{p_j p_j}$, $c_{g_j g_j}$ will have a 1 added and the offdiagonal coefficients represented by products such as μf_i will also have a 1 added. The symmetry of the coefficients allows storing only the diagonal elements and one side of the off-diagonal elements. Optimum strategies for summing and storing the coefficients will depend on the data set and computing equipment.

After the least squares coefficients are accumulated, the least squares equations can be modified to make them into mixed model equations by adding the ratios of residual to other variances to the proper parts of C. The mixed model or least squares equations can then be solved. One method of solving a large number of equations such as mixed model equations is by iteration (the method of successive improvement).

SOLVING EQUATIONS

In many situations, the number of least squares or mixed model equations is so large that an inverse of the coefficient matrix cannot be computed even though many strategies have been developed to reduce the number of equations. If prediction error variances are not needed, solutions can be obtained by iteration. The most efficient computing strategy will depend on the model, the amount of data, and computing equipment. The augmented mixed model equations for the animal model are especially well-suited to innovative computing strategies. Nevertheless, the basic principle of Gauss-Seidel iteration will be demonstrated with three equations. Other methods of iteration follow a similar pattern.

Let the equations be Cs = r, where C is the symmetric matrix of coefficients, s is the vector of solutions and r is the vector of right-hand sides. Then for three equations:

$$c_{11}s_1 + c_{12}s_2 + c_{13}s_3 = r_1$$
$$c_{21}s_1 + c_{22}s_2 + c_{23}s_3 = r_2$$
$$c_{31}s_1 + c_{32}s_2 + c_{33}s_3 = r_3 .$$

The method of Gauss-Seidel iteration will be illustrated for this set of three equations where the c_{ij} are known numerical coefficients of the unknown solutions, the s_i, and the r_i are known numerical values in the RHS vector. The steps to obtain solutions by iteration are as follows.

ITERATION

Step 1: To begin the iteration, guess a set of initial solutions for s; s_1^o, s_2^o, s_3^o. The starting values should approximate the expected values of the solutions.

Step 2: The basic step for each equation is to solve for that solution after substituting solutions from the same or the previous round of iteration for the other solutions.

Round 1

i) Solve for s_1 with s_2^o and s_3^o:

$$s_1^1 \leftarrow (1/c_{11})[r_1 - c_{12}s_2^o - c_{13}s_3^o]$$

Replace the previous solution for s_1 with s_1^1.

ii) Solve for s_2 with s_1^1 and s_3^o:

$$s_2^1 \leftarrow (1/c_{22})[r_2 - c_{21}s_1^1 - c_{23}s_3^o]$$

Replace the previous solution for s_2 with s_2^1.

iii) Solve for s_3 with s_1^1 and s_2^1:

$$s_3^1 \leftarrow (1/c_{33})[r_3 - c_{31}s_1^1 - c_{32}s_2^1]$$

Replace the previous solution for s_3 with s_3^1.

Round 2 to round **n**

i) $s_1^n \leftarrow (1/c_{11})[r_1 - c_{12}s_2^{n-1} - c_{13}s_3^{n-1}]$

ii) $s_2^n \leftarrow (1/c_{22})[r_2 - c_{21}s_1^n - c_{23}s_3^{n-1}]$

iii) $s_3^n \leftarrow (1/c_{33})[r_3 - c_{31}s_1^n - c_{32}s_2^n]$

Note that the most current estimates in s are used to update each solution from its equation. For example, the Jacobi method does not update s until the end of the round.

An equivalent expression for s_i^n is:

$$s_i^n \leftarrow s_i^{n-1} + (1/c_{ii})[r_i - \sum_{j=1}^{i-1} c_{ij}s_j^n - c_{ii}s_i^{n-1} - \sum_{j=i+1} c_{ij}s_j^{n-1}] \quad .$$

This expression requires an extra multiplication and two extra additions per equation per round. The advantages may outweigh the extra arithmetic. Solving equations by iteration requires a rule for stopping the iteration. Such a rule can be based on the expression in square brackets on the right which is zero when the solutions are exact. Thus, this difference between the right-hand side and the right-hand side regenerated from estimates in the most recent round of iteration is often the basis for the stopping criterion. One such criterion is $(\Sigma e_i^2)^{.5} / (\Sigma r_i^2)^{.5}$ where

$$e_i = [r_i - \sum_{j=1}^{i-1} c_{ij}s_j^n - c_{ii}s_i^{n-1} - \sum_{j=i+1} c_{ij}s_j^{n-1}]$$

Dividing by $(\Sigma r_i^2)^{.5}$ scales the solutions for the trait being analyzed. Iteration is stopped when at the end of a round the stopping criterion is less than a pre-set value, e.g., .01 or .001.

Another advantage of the second form is that a modification of Gauss-Seidel iteration called successive-over-relaxation (SOR) is easy to implement:

$$s_i^n \leftarrow s_i^{n-1} + (w/c_{ii})[r_i - \sum_{j=1}^{i-1} c_{ij}s_j^n - c_{ii}s_i^{n-1} - \sum_{j=i+1} c_{ij}s_j^{n-1}]$$

where w is the relaxation factor. A relaxation factor larger than 1 but less than 2 is likely to result in faster convergence than with Gauss-Seidel iteration (w=1). One difficulty with SOR is how to determine the optimum **w** before beginning to iterate.

CHAPTER 33

MODELS FOR CROSSBREEDING

Whether a chapter on crossbreeding should be in the part on selection index or among the chapters on mixed model methods poses a problem. In most introductory animal genetics textbooks the description of crossbreeding implies all effects in the model are constants depending on effects of breeds and heterosis effects. That approach fits completely neither Part I nor Part II. A true model for prediction of breeding values from crossbred data, however, also includes the genetic deviations of individual animals from the breed and heterosis constants. Because the breed and heterosis constants usually must be estimated from the same data used to predict the deviations, then the appropriate model is a mixed model including the breed and heterosis constants as well as the genetic deviations. This discussion, therefore, will assume that breed and heterosis constants have been estimated and genetic deviations predicted with mixed model equations for an animal model that might include both direct and maternal genetic effects as described in Part II. One difficulty is that choice of constraints might not be easy. With some designs complete confounding may occur between breed and specific heterosis effects. The goal of crossbreeding generally is to combine breeds to maximize the breed and heterosis effects. The goal of this chapter will be to show how to calculate combined breed and heterosis effects from known or estimated breed and specific heterosis effects. The predicted genetic deviations for direct

and maternal effects can be added to predict direct and maternal performance of specified matings. Most introductory animal breeding textbooks describe typical mating plans for crossbreeding, such as two- and three-way rotations, terminal crosses, grading-up, and creation of composite or synthetic breeds. This chapter will not duplicate those tables. Instead the models and algorithms for computing the coefficients for the effects included in specific breed combinations will be developed.

Crossbreeding theory can easily be put into the simple animal model. In fact, the theory needed is quite minor but the notation is horrendous!! The simplified model (subtracting out MANAGEMENT effects, etc.) will be used. A simplifying assumption made in the first part of this discussion, is that the average for the breed represents every animal of the breed, both for additive direct and maternal effects and also for heterotic direct and maternal effects. The breeding value for an animal, however, is its breed constants plus direct and maternal genetic deviations from those constants which will be added later in the discussion.

HETEROSIS

Heterosis arises from the crossing of breeds or inbred lines. The measure of *heterosis* is the difference in performance between the cross and the average of the parent breeds. Heterosis may be positive, negative or nil. Heterosis also can affect maternal traits.

Several genetic mechanisms can explain heterosis. The most usual explanation involves dominance effects and the assumption that at some genetic loci, different genes have become fixed for different breeds. Fixation means that all genes at that chromosome

location are identical. *Dominance* means that at a chromosome location with two possible types of genes, one of the genes will dominate the other. An example is the gene for black, B, in cattle. The other gene at the same chromosome location is for red, b. The three possible genotypes and their phenotypes are:

Genotype	Phenotype
BB	Black
Bb or bB	Black
bb	Red

Either a pair of B genes or a single B gene results in black. In genetic jargon, the gene for black, B, dominates the gene for red, b, i.e., B is dominant to b. Conversely, b is said to be *recessive* to B, i.e., hides when B is present.

The important point for quantitative effects is that with desirable dominant genes, either one (heterozygote) or two (homozygote) of the desirable genes gives the same response. In rare cases for some traits the effect of the single dose, Bb, may exceed that of the double dose, BB. Then the term *over-dominance* is used. The simplified example for just four loci will involve only dominance. The idea is that desirable dominant genes at some chromosome locations have become fixed in one breed but that other, less desirable, recessive genes have been fixed at those chromosome locations in the other breed. At other chromosome locations, the reverse is true. As an example, assume four loci have become fixed in breeds I and II for the "A", "B", "C", and "D" loci. Upper case letters will indicate dominant alleles and lower case letters, the recessive alleles. Breed I is fixed as (AA, bb, cc, DD) and breed II is fixed as (aa, BB, CC, DD). The desirable genes are fixed at "A" and "D" for breed I and at "B", "C", and "D" for breed II. The genotype of the cross

will be (Aa, Bb, Cc, DD). Now assume that with at least one upper case gene, the effect at each locus is +2; and that with no upper case genes, the effect is 0.

Breed	Genes at chromosome location "A"	"B"	"C"	"D"	Value of genes at chromosome location "A"	"B"	"C"	"D"	Sum
I	AA	bb	cc	DD	2	0	0	2	4
II	aa	BB	CC	DD	0	2	2	2	6
Cross	Aa	Bb	Cc	DD	2	2	2	2	8

Heterosis = Cross - average (I and II)

Heterosis = 8 - (4+6)/2 = 3

In this example, the cross performance exceeds performance of both parents and the average of the parents by 60%. Heterosis is defined as a percentage of the average of the parent breeds:

$$\text{Heterosis } (\%) = \left(\frac{\text{Cross - average of parents}}{\text{average of parents}}\right) \times 100 \quad .$$

In the example

$$\text{Heterosis } (\%) = \left[\frac{8 - (4+6)/2}{(4+6)/2}\right] \times 100 = 60\% \quad .$$

The example shows that the D locus does not contribute to heterosis because both breeds are fixed for the desirable dominant gene, i.e., both are DD. For many breeds, most gene locations probably do not have dominant alleles. Desirable dominant alleles may also be the same at many gene locations of most breeds. A few gene locations, such as "A", "B", and "C", where different good dominant genes have become fixed in two breeds with recessive genes fixed in the other breed, however, can lead to heterosis percentages of 3 to 15%, which are common for crosses among many breeds for many traits.

MODELS FOR BREED CROSSES

The notation for models with breed crosses will be for upper case letters to indicate a breed. The breed constant for each animal of a breed represents the breed average. Let $\overline{A}\overline{A}$ be the average of pure breeding of breed A. Similarly, $\overline{A}\overline{B}$ will represent the average of progeny of crosses from mating sires of breed A to dams of breed B. The usual convention in describing a breed cross is to write the sire breed first. In the model for cross performance, no heterosis is allowed within a breed, e.g., $\overline{A}\overline{A}$ does not contain any heterosis effects.

The notation used for describing cross performance will be illustrated with a two-way cross; records of progeny of sires of breed A and dams of breed B, with expected performance, $\overline{A}\overline{B}$. **DIR** in front of symbols will represent direct genetic effects and **MAT** in front of symbols will represent maternal genetic effects.

The model for $\overline{A}\overline{B}$ is:

$$\overline{A}\overline{B} = \mathbf{DIR}[AA/2 + BB/2 + H(AB)] + \mathbf{MAT}[BB].$$

The **DIR**[AA/2] and **DIR**[BB/2] represent the additive genetic (breeding value) contributions of the parents to their progeny. The extra effect of heterosis from a breed A by breed B cross is denoted as H(AB). The first three terms correspond to the direct effects of the genotype of the animal model. A more complete model as described later includes those three terms as contributing to genetic means and also includes deviations from those means representing individual animal genetic differences from breed averages. The fourth term, **MAT**(BB), represents the breed average genetic *maternal* effect; in this case, the maternal effect of breed B, the dam breed.

MATERNAL HETEROSIS

In many cases, an advantage of crossbreeding is due to heterosis for maternal effects. Crossbred dams can exhibit maternal heterosis. For such crosses, maternal heterosis must be incorporated into the model. Suppose the dams are from a cross of breeds C and D. The model for maternal effects is:

$$\mathbf{MAT}[(CD)] = \mathbf{MAT}[CC/2 + DD/2 + H(CD)].$$

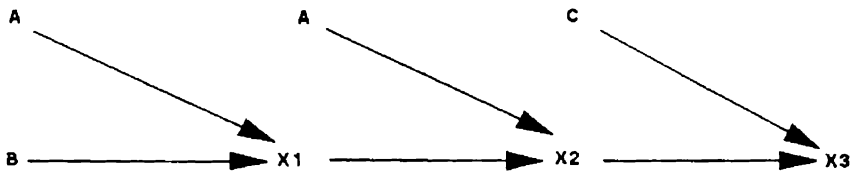
This model is the same as for direct effects, except that **MAT** has been put in front to indicate that CC/2 and DD/2 refer to average maternal effects of breeds C and D transmitted to their cross daughters, CD, and H(CD) refers to the maternal heterosis.

With more than two breeds in a cross, the model becomes progressively more complicated. The additive direct effects of the breeds are weighted by the proportions of genes they contribute to the final cross. Similarly, breed contributions to additive maternal effects are based on the cross animals used as dams. Somewhat more difficult to calculate are fractions of specific heterosis effects contributed by the various breed combinations.

COMPUTING BREED CONTRIBUTIONS TO CROSSES

Two simple computing procedures can be used to calculate the correct fractions for breed direct effects and breed by breed heterosis effects, no matter how complicated the crossbreeding breeding plan is. Again, an example will describe intuitively how to do these simple computations.

Suppose the following crosses among breeds A, B, and C are involved:



Direct contributions of each breed to the crosses can be calculated by counting paths, but an easier way is to set up a simple table and remember a parent contributes one-half of its genes to its progeny. Simply treat each breed as a parent and use a modification of the tabular method of calculating relationships. The three steps are:

1) Along the top of the table write the breeds and combinations with the parents of each cross listed above the combinations in the order they arise. Along the left side, write the breeds contributing to the breeding plan.

	A	B	C	A-B X_1	A-X_1 X_2	C-X_2 X_3
A						
B						
C						

2) Pretend relationships are to be calculated; this step corresponds to calculation of fractions of genes in each cross that trace back to parent breeds.

	A	B	C	A-B X_1	A-X_1 X_2	C-X_2 X_3
A	1					
B		1				
C			1			

3) Calculate entries to the right of the pure breeds as one-half the entry for the first parent in that row plus one-half the entry for the second parent in that row.

For example, the (A, X_1) entry is one-half of the (A,A) entry of 1 plus one-half of the (A,B) entry of 0: the (A, X_1) entry = (1)/2 + (0)/2 = 1/2.

The (A, X_2) entry is one-half of the (A,A) entry of 1 plus one-half of the (A, X_1) entry of 1/2 calculated above: the (A, X_2) entry = (1)/2 + (1/2)/2 = 3/4.

The completed table is:

	A	B	C	A-B X_1	A-X_1 X_2	C-X_2 X_3
A	1			1/2	3/4	3/8
B		1		1/2	1/4	1/8
C			1	0	0	1/2

The fractions in each column represent the breed contributions to that cross. For example, of the genes for X_3, 3/8 are expected to come from breed A, 1/8 from breed B and 1/2 from breed C. No surprises, as these could have been calculated by rough reasoning. The advantage of the table comes with more complicated crosses. Suppose X_2 by X_3 matings are made to produce X_4? All that is needed is to add a column for X_4 and follow the rule of the sum of one-half of the entries for each of the parents, X_2 and X_3:

	A	B	C	A-B X_1	A-X_1 X_2	C-X_2 X_3	X_2-X_3 X_4
A	1			1/2	3/4	3/8	9/16
B		1		1/2	1/4	1/8	3/16
C			1	0	0	1/2	4/16
Sum	1	1	1	1	1	1	1

The columns always sum to 1 or 100%. The direct additive effects and maternal additive effects for X_4 are easily modeled from the table with X_3 being the dam group:

$\bar{X}_4\bar{X}_4$ = **DIR**[(9/16)AA + (3/16)BB + (4/16)CC] + **MAT**[(3/8)AA + (1/8)BB + (1/2)CC]

COMPUTING POTENTIAL HETEROSIS FROM CROSSES

The contributions to direct and maternal heterosis are still needed. This time the computing trick is intuitively a little more fuzzy, but it works! The computation is based on multiplying the gametic arrays (where genes come from and their proportions) of the two parent groups. The result shows how much heterozygosity and potentially how much heterosis to expect.

For example start with a cross of A by B. Breed A contributes only A genes so the gametic array is 1A. Similarly, the gametic array for breed B is 1B. Now multiply and as expected:

$$\text{for A x B} \longrightarrow X_1 \text{ ; (1A) x (1B) = 1 AB.}$$

All of the X_1 cross are AB; the heterozygous condition is 100%.

For the cross of A by X_1, the gametic array for X_1 can be obtained from the column of the table for X_1. The fractions, 1/2 for A and 1/2 for B, represent the origins of the genes for X_1 and make up the gametic array for X_1: [(1/2)A + (1/2)B]. Thus:

$$\text{for A x } X_1 \longrightarrow X_2\text{; (1A) x [(1/2)A + (1/2)B] = (1/2)AA + (1/2)AB.}$$

By inspection, 50% of the cross is heterozygous, AB. The backcross to A has resulted in loss of one-half of the potential heterosis.

What about the cross, C by X_2?

For C x $X_2 \longrightarrow X_3$; (1C)[(3/4)A + (1/4)B] = (3/4)AC + (1/4)BC.

The elements of gametic array, 3/4 and 1/4, come from the column for X_2! Inspection shows heterozygosity is 100%; 75% of the heterosis is expected from AC and 25% of the heterosis is expected from BC. For practice try the X_2 by X_3 cross? The potential heterosis is 6/32 of AB heterosis, 12/32 of AC heterosis and 4/32 of BC heterosis. Note that the specific heterosis effects may not be equal, i. e., H(AB) may not equal H(AC) may not equal H(BC).

HETEROSIS IN THE CROSSBRED MODEL

Now heterosis can be added to the model. To make the computations a little easier, the simple cross of A (sire breed) by X_1 (crossbred dams) will be examined.

The fractions of direct heterosis effects come from the A by X_1 gametic arrays:

DIR[(1/2)H(AA) + (1/2)H(AB)].

By definition, heterosis within a breed does not exist, H(AA) = 0 and direct heterosis is:

DIR[(1/2)H(AB)].

The maternal effects are contributed by X_1: thus, heterosis comes from the gametic arrays that produced X_1; (1A) x (1B) = 1 AB .

The maternal heterosis is 100% of **MAT**[H(AB)].

The full model for $\bar{X}_2\bar{X}_2$ becomes:

$\bar{X}_2\bar{X}_2$ = **DIR**[(3/4)AA + (1/4)BB + (1/2)H(AB)] + **MAT**[AA/2 + BB/2 + H(AB)] .

This cross exhibits 50% of potential direct heterosis, **DIR**[H(AB)] and 100% of potential maternal heterosis **MAT**[H(AB)].

SYNTHETIC OR COMPOSITE BREEDS

Often breeds are crossed in various ways and at a certain stage, the crossbred animals are mated *inter se* (randomly among themselves) to develop after a few generations a *synthetic* or *composite* breed. The preceding model can be used to predict the average performance of a synthetic breed from the specific breed and heterosis effects. Suppose that X_3 is chosen as the group to begin *inter se* mating: C x [A x (AB)]. Contribution to both additive direct and maternal effects (X_3's will also be the mothers) will be determined by the breed contributions of A, B, and C shown in the column for X_3:

DIR[(3/8)AA + (1/8)BB + (1/2)CC] for direct effects and

MAT[(3/8)AA + (1/8)BB + (1/2)CC] for maternal effects.

Potential heterosis is computed from the gametic arrays:

[(3/8)A + (1/8)B + (1/2)C] x [(3/8)A + (1/8)B + (1/2)C]

= (9/64)AA + (1/64)BB + (16/64)CC + (6/64)AB + (24/64)AC + (8/64)BC.

The expected model for $\bar{X}_3\bar{X}_3$ is:

DIR[(3/8)AA + (1/8)BB + (1/2)CC
+ (3/32)H(AB) + (12/32)H(AC) + (4/32)H(BC)]
+ **MAT**[(3/8)AA + (1/8)BB + (1/2)CC
+ (3/32)H(AB) + (12/32)H(AC) + (4/32)H(BC)].

Obviously this model is symbolic. What is needed are the numerical values associated with the specific symbols. Research at various experiment stations such as the USDA Meat Animal Research Center (MARC) has had the goal of estimating those effects.

RETENTION OF HETEROZYGOSITY

The retention of heterozygosity in *inter se* mating of the X_3 in the previous section is:

6/64 of H(AB), 24/64 of H(AC) and 8/64 of H(BC).

The loss of heterozygosity as a fraction of complete heterozygosity is:

$$9/64 + 1/64 + 16/64 = 26/64.$$

The formula for retention of heterozygosity and potential heterosis, if all heterosis effects are equal, is:

$$1 - p_A^2 - p_B^2 - p_C^2$$

where p_A, p_B, and p_C are the fractions of genes from breeds A, B, and C in the cross used for *inter se* mating. The general formula with n breeds involved is:

$$1 - \sum_{i}^{n} p_i^2$$

when p_i is the fraction of genes contributed by breed i.

COMPLEMENTARITY

Complementarity is a potential benefit of crossbreeding that is not as glamorous as heterosis but which is often very important to the success of crossbreeding programs or of synthetic breeds. The term, complementarity, describes the concept. Weaknesses of one breed are improved by strengths of other breeds. A more exciting name for the result was coined by Moav -- *economic heterosis*. Economic heterosis can occur without heterosis of any single economic trait, although heterosis usually contributes to economic heterosis. Often positive complementarity arises because of a multiplier trait, e.g., percent calf crop, live pigs weaned per litter, fat percentage for dairy cows. Reproduction and viability traits

are multiplier traits and also often exhibit positive heterosis. As a simple example, suppose two breeds present the following profile.

Breed	Percent calf crop	Weaning weight	Weight sold per cow
A	90	500	450
B	70	600	420
AB	80	550	440

In this case, the economic heterosis is 440 - (450 + 420)/2 = 5 lb. Economic heterosis is only about 1% but in this example neither percent calf crop nor weaning weight exhibit heterosis. As with single trait heterosis, however, economic heterosis may be negative.

PREDICTION OF PROGENY RESPONSE AND BREEDING VALUE

The goal of this chapter was to predict performance of progeny from the mating of a particular sire of some breed or breedcross and a dam of some breed or breedcross. The difficult part of the theory has now been discussed. What is left to do is put together the breed and heterosis effects with the direct and maternal genetic deviations. The steps are:

1) estimate breed constants for direct and maternal effects and specific breed by breed direct and maternal heterosis effects (not an easy task in many cases),

2) predict direct and maternal genetic deviations for potential sires and dams either jointly with estimating the breed and heterosis constants or after adjusting records for those pre-estimated constants.

Expected progeny performance will depend on breed composition of the two parents as well as their predicted genetic deviations for direct and maternal effects; a_S and m_S for the sire and a_D and m_D for the dam. The equation is for a single trait but in most cases economic heterosis (complementarity over all traits) should be considered. For example, assume the potential sire is of breed A and the potential dam is a cross of breeds B and C. Expected progeny performance is the sum of direct genetic breed and heterosis constants and genetic deviations associated with the sire and dam plus maternal genetic effects associated with the breed composition and genetic deviation of the dam. For the example:

$$\mathbf{DIR}[AA/2 + BB/4 + CC/4 + (1/2)H(AB) + (1/2)H(AC)] + (a_S + a_D)/2$$
$$+ \mathbf{MAT}[BB/2 + CC/2 + H(BC)] + m_D \quad .$$

The expected progeny performances for all potential sires when mated to all potential dams could be compared to determine the matings that would be expected to maximize expected progeny performance or economic heterosis.

The breeder may be interested in the breeding value of the potential progeny rather than the progeny performance. The predicted breeding value of the progeny, however, does not depend on the heterosis effects which cannot be transmitted directly to its descendents. One-half of the breed effects are transmitted to the next generation and thus are included with the genetic deviations in the predicted breeding values.

The direct breeding value for the example is:

$$\mathbf{DIR}[AA/2 + BB/4 + CC/4] + (a_S + a_D)/2.$$

The maternal breeding value for the example is:

$$\mathbf{MAT}[AA/2 + BB/4 + CC/4] + (m_S + m_D)/2 \quad .$$

The maternal breeding value of female progeny would, in turn, be expected to be expressed in the performance of their progeny.

The expected breeding value of progeny of a sire and dam is the average of the breeding values of the sire and dam with breed effects included in breeding values of both the progeny and its sire and dam. Progeny of an A by (B by C) mating would probably create some heterosis when they become parents. The potential heterosis should not be included as part of their breeding values because the heterosis effects will depend on the breed composition of the potential mates.

LAST THOUGHTS ABOUT CROSSBREEDING

The management aspects can be difficult.

Choice of breeds is important.

Heterosis and complementarity depend on breeds available. All breed crosses may not result in equal heterosis effects or average complementarity.

The whole package -- specific direct and maternal additive genetic values, specific heterotic effects for direct and maternal effects of all traits, complementarity, management costs, and markets -- must be put together properly for crossbreeding programs to be successful.

Progeny performance can be considered for short-term planning but progeny breeding values must be considered for long term breeding goals.

CHAPTER 34

FLOW CHART FOR MIXED MODEL EQUATIONS

I. Determine the model

A. Fixed factors, e.g., age, management group

B. Random factors (other than **G**, **G**/2, **PE**)

C. **G** + **E**; $\sigma_g^2 + \sigma_e^2 = \sigma_P^2$; $\sigma_y^2 = \sigma_P^2$ + sum of other σ_r^2's

1. Animal model:

G; **w** = **E**,

$\sigma_g^2 = h^2\sigma_P^2$, $\sigma_w^2 = (1 - h^2)\sigma_P^2$

2. Repeated records animal model:

G, **PE**; **w** = **TE**,

$\sigma_g^2 = h^2\sigma_P^2$, $\sigma_{pe}^2 = (r - h^2)\sigma_P^2$, $\sigma_w^2 = (1 - r)\sigma_P^2$

3. Sire model:

s = **G**/2; **w** = other **G** + **E**,

$\sigma_s^2 = h^2\sigma_P^2/4$, $\sigma_w^2 = (1 - h^2/4)\sigma_P^2$

II. Create Least Squares Equations from (rules: sums → model, ^'s)

III. Modifications for MME (animals or sires **unrelated**)

A. Add σ_w^2/σ_r^2 to diagonal coefficients, other random factors

B. Animals with records, unrelated

1. Add σ_w^2/σ_g^2 to diagonal coefficients of the **G** equations

2. If repeated records, add σ_w^2/σ_{pe}^2 to diagonals of **PE** equations

C. Sires with progeny with records, unrelated sires

1. Add σ_w^2/σ_s^2 to diagonal coefficients of the sire equations

IV. Modifications for MME (animals or sires **related**)

A. Add σ_w^2/σ_r^2 to diagonal coefficients of other random factors.

B. Animals with records related through, **A** matrix.

1. If repeated records, add σ_w^2/σ_{pe}^2 to diagonal coefficients of the **PE** equations.
2. Add A^{-1} (σ_w^2/σ_g^2) to the g x g block of coefficients.

or

3. Calculate A_+^{-1} directly by the Henderson rules (noninbred).
   a) Include base animals with no records if related to more than one animal with records.
   b) Augment equations to include animals with no records.
      i) sum = 0; ii) no model; iii) tied by A_+^{-1} (σ_w^2/σ_g^2)
   c) Jointly predict **G**; animals with records and base animals with no records.

C. Sires with progeny with records. **A** is matrix of relationships.

1. Add A^{-1} (σ_w^2/σ_s^2) to s x s block of coefficients.

or

2. Calculate A_+^{-1} directly by rules for (noninbred) sire or for sire and maternal grandsire.
   a) Include base animals with no progeny with records if related to more than one sire with progeny.
   b) Augment equations to include animals with no progeny records.
      i) sum = 0; ii) no model; iii) tied by A_+^{-1} (σ_w^2/σ_s^2)
   b) Jointly predict **G**/2 for sires with progeny records and relatives with no progeny records.

PART THREE

ESTIMATING GENETIC PARAMETERS USING SIMPLE STATISTICAL MODELS

The following chapters cover simple procedures used in estimating repeatability, heritability and genetic, environmental, and phenotypic correlations. In most cases, the complete statistical model is given including the expectations or average values of all relevant combinations of the observations, as well as a worked example. Methods of simulating the models are described in the last two chapters.

CHAPTER 35

SUMMATION AND DOT NOTATION

SUMMATION NOTATION

Summation and dot notation are used in describing the computations for estimating genetic parameters. A Σ indicates summation over what follows as the subscripts vary by 1 from the lower limit of summation to the upper limit. For example,

$$\sum_{i=1}^{3} X_i = X_1 + X_2 + X_3 .$$

Similarly,

$$\sum_{j=1}^{4} j = 1 + 2 + 3 + 4 .$$

If there is no subscript in what follows, the quantity is simply repeated the number of times the limits of summation indicate:

$$\sum_{k=1}^{3} c = c + c + c = 3\,c .$$

If the summation limits are not given, the limits are usually obvious.

DOT NOTATION

A dot in place of a subscript signifies summation over that subscript. As an example of this notation, consider a set of observations denoted as P_{ij}. Let $i = 1, ..., 3$ and $j = 1, ..., n_i$ where $n_1 = 2$, $n_2 = 1$, $n_3 = 4$. This arrangement could correspond to three animals (the first subscript) where the first animal has $n_1 = 2$ records, the second animal $n_2 = 1$ record, and the third animal $n_3 = 4$ records. The various quantities to be written out in this example correspond to quantities which will be used later in estimating repeatability. Also keep in mind that each symbol corresponds to a number in an actual analysis, e.g., P_{11} is the first record on the first animal.

$$\text{Then } \sum_{i=1}^{3} \sum_{j=1}^{n_i} P_{ij} = P.. = P_{11} + P_{12} + P_{21} + P_{31} + P_{32} + P_{33} + P_{34}.$$

The subscript corresponding to the innermost Σ usually varies first:

$$\sum_{i=1}^{3} \sum_{j=1}^{n_i} P_{ij} = \sum_{i=1}^{3} P_{i.} = P_{1.} + P_{2.} + P_{3.} = P..$$

$$P_{1.} = \sum_{j=1}^{n_1=2} P_{1j} = P_{11} + P_{12}$$

$$P_{2.} = \sum_{j=1}^{n_2=1} P_{2j} = P_{21}$$

$$P_{3.} = \sum_{j=1}^{n_3=4} P_{3j} = P_{31} + P_{32} + P_{33} + P_{34}$$

$$P.. = P_{1.} + P_{2.} + P_{3.} = P_{11} + P_{12} + P_{21} + P_{31} + P_{32} + P_{33} + P_{34} = \sum_{i=1}^{3} \sum_{j=1}^{n_i} P_{ij} .$$

The dot indicates that the corresponding summation has been finished before doing other operations such as further summation or squaring.

Some examples with sums of squares used for estimating repeatability are as follows:

$$\sum_{i=1}^{3} \frac{\left(\sum_{j=1}^{n_i} P_{ij}\right)^2}{n_i} = \sum_{i=1}^{3} \frac{P_{i.}^2}{n_i} = \frac{P_{1.}^2}{n_1} + \frac{P_{2.}^2}{n_2} + \frac{P_{3.}^2}{n_3} = \frac{(P_{11}+P_{12})^2}{2} + \frac{(P_{21})^2}{1} + \frac{(P_{31}+P_{32}+P_{33}+P_{34})^2}{4}$$

$$\frac{\left(\sum_{i=1}^{3}\sum_{j=1}^{n_i} P_{ij}\right)^2}{\sum_{i=1}^{3} n_i} = \frac{P_{..}^2}{n.} = \frac{(P_{1.}+P_{2.}+P_{3.})^2}{n_1+n_2+n_3} = \frac{(P_{11}+P_{12}+P_{21}+P_{31}+P_{32}+P_{33}+P_{34})^2}{(1+1)+(1)+(1+1+1+1)}$$

$$\sum_{i=1}^{3}\sum_{j=1}^{n_i} P_{ij}^2 = P_{11}^2 + P_{12}^2 + P_{21}^2 + P_{31}^2 + P_{32}^2 + P_{33}^2 + P_{34}^2 .$$

Note that there is nothing comparable to the dot notation for sums of squared quantities.

The same procedure applies to functions of the n's.

$$n. = n_1 + n_2 + n_3 = 2 + 1 + 4 = 7 .$$

$$\sum_{i=1}^{3} n_i^2 = n_1^2 + n_2^2 + n_3^2 = 2^2 + 1^2 + 4^2 = 21 .$$

$$n.^2 = (n_1 + n_2 + n_3)^2 = (2 + 1 + 4)^2 = 49 .$$

$$\left(\sum_{i=1}^{3} n_i^2\right) / n. = 21/7 = 3 .$$

CHAPTER 36

EXPECTED VALUES

The usual technique in estimating heritability, repeatability, and various correlations is to compute certain functions of the data (usually sums of squares or sums of cross products corresponding to those computed for balanced analyses of variance) and equate the functions to their expected or average values. The sums of squares or crossproducts to be considered are the usual ones. An appreciation of how to find expected values, however, is needed. More complicated models including fixed effects will not be considered. Only models with all effects random will be discussed. The general method was described by Crump (1946, 1951) and Henderson (1953).

The symbol most often used for the expected or average value of some expression, involving constants and variables, is E( ). Expected values of most expressions used in estimating genetic parameters are relatively easy to find if the following definitions are remembered.

DEFINITIONS

Let c = a constant; x_i = a variable from some distribution of trait X with mean, μ_x, and variance, σ_x^2; and y_i = a variable from some distribution of trait Y with mean μ_y, variance σ_y^2, and covariance with x_i, σ_{xy}.

Definition 1: E(c) = c. The average value of a constant is that constant. Similarly $E(c^2) = c^2$.

Definition 2: $E(x_i) = \mu_x$. The average of all possible values of variable X is its average or mean, μ_x.

Definition 3: $E(cx_i) = c\ E(x_i) = c\ \mu_x$. The average of all possible values of a variable times a constant is the constant times the expected value of the variable. The principle is that in expressions involving a constant the constant can be taken outside the expectation operation.

Definition 4: $E(x_i+y_j) = E(x_i) + E(y_j) = \mu_x + \mu_y$. The expectation of a sum can be taken as the sum of the expectations of the parts. The principle is that expectations of parts of a function can be done separately and then added together.

Definition 5: $E(x_i - \mu_x)^2 = \sigma_x^2$. By definition the variance of a variable X, σ_x^2, is the average squared deviation of the variable from its mean.

Thus, $E(x_i^2) = \sigma_x^2 + \mu_x^2$ which follows directly from definition 5. Expand the equation for definition 5 and take the expectations of its parts:

$$\sigma_x^2 = E(x_i - \mu_x)^2 = E(x_i^2 - 2x_i\mu_x + \mu_x^2),$$

$$= E(x_i^2) - E(2\ \mu_x\ x_i) + E(\mu_x^2) \text{ from (4)},$$

$$= E(x_i^2) - 2\ \mu_x\ E(x_i) + \mu_x^2 \text{ from (1) and (3)},$$

$$= E(x_i^2) - 2\ \mu_x\ \mu_x + \mu_x^2,$$

$$= E(x_i^2) - \mu_x^2 \quad \text{so that}$$

$$E(x_i^2) = \sigma_x^2 + \mu_x^2 \ .$$

Note when $\mu_x = 0$ that $E(x_i^2) = \sigma_x^2$.

Definition 6: $E[(x_i - \mu_x)(y_i - \mu_y)] = \sigma_{xy}$. By definition the covariance between variables X and Y, σ_{xy}, is the average of the products of their deviations from their means.

Thus, $E(x_iy_i) = \sigma_{xy} + \mu_x\mu_y$ which follows from definition 6. Expand the equation for definition 6 and take the expectations of its parts:

$\sigma_{xy} = E[(x_i - \mu_x)(y_i - \mu_y)]$,

$= E[x_iy_i - \mu_xy_i - \mu_yx_i + \mu_x\mu_y]$,

$= E(x_iy_i) - \mu_xE(y_i) - \mu_yE(x_i) + \mu_x\mu_y$, from (1) and (3),

$= E(x_iy_i) - \mu_x\mu_y - \mu_y\mu_x + \mu_x\mu_y$,

$= E(x_iy_i) - \mu_x\mu_y$ so that

$E(x_iy_i) = \sigma_{xy} + \mu_x\mu_y$. When either μ_x or $\mu_y = 0$, then $E(x_iy_i) = \sigma_{xy}$.

The general procedure for applying these definitions to find the expected values of more complicated sums of squares and products of variables is to use the following steps.

Step 1. Substitute elements of the model into the function.

Step 2. Expand the function in terms of the model.

Step 3. Find the expected value of each term of the function.

Step 4. The expected value of the function will be the sum of the expected values of the individual terms.

EXAMPLE

Let $P_{ij} = \mu + A_i + E_{ij}$ where P_{ij} is an observation (variable) on the j^{th} record in the i^{th} class, μ is a constant, A_i is a variable with $\mu_A = 0$ and variance, σ^2_A, E_{ij} is a variable with $\mu_E = 0$ and variance σ^2_E, and the covariance between any two A's, any two E's or any A and E is zero.

The expected value of any observation:

$$E(P_{ij}) = E(\mu + A_i + E_{ij}) = E(\mu) + E(A_i) + E(E_{ij})$$
$$= \mu + 0 + 0$$
$$= \mu .$$

The expected value of any observation squared:

$$E(P_{ij}^2) = E[(\mu+A_i+E_{ij})^2] = E(\mu^2+A_i^2+E_{ij}^2+2\mu A_i+2\mu E_{ij}+2A_iE_{ij})$$
$$= E(\mu^2) + E(A_i^2) + E(E_{ij}^2) + E(2\mu A_i) + E(2\mu E_{ij}) + E(2A_iE_{ij})$$
$$= \mu^2 + \sigma_A^2 + \sigma_E^2 + 2\mu E(A_i) + 2\mu E(E_{ij}) + 2E(A_iE_{ij})$$
$$= \mu^2 + \sigma_A^2 + \sigma_E^2 \qquad \text{because } E(A_i) = 0,\ E(E_{ij}) = 0,$$
$$\text{and } E(A_iE_{ij}) = \sigma_{AE} = 0 .$$

The expected value of the product of observations in the same class:

$$E(P_{ij}P_{ij'}) = E[(\mu + A_i + E_{ij})(\mu + A_i + E_{ij'})] \text{ for } (j' \neq j)$$
$$= E(\mu^2+\mu A_i+\mu E_{ij'}+\mu A_i+A_i^2+A_iE_{ij'}+\mu E_{ij}+A_iE_{ij}+E_{ij}E_{ij'})$$
$$= \mu^2 + 0 + 0 + 0 + \sigma_A^2 + 0 + 0 + 0 + 0$$
$$= \mu^2 + \sigma_A^2, \text{ because } E(A_iE_{ij'}) = 0 = E(A_iE_{ij}) \text{ since } \sigma_{AE} = 0 \text{ and}$$
$$E(E_{ij}E_{ij'}) = 0 \text{ since } \sigma_{E_{ij}E_{ij'}} = 0 .$$

The expected value of the product of observations in different classes:

$$E(P_{ij}P_{i'j'}) = E[(\mu + A_i + E_{ij})(\mu + A_{i'} + E_{i'j'})] \quad (i' \neq i \text{ and } j=j' \text{ or } j \neq j')$$
$$= E[\mu^2+\mu A_{i'}+\mu E_{i'j'}+\mu A_i+A_iA_{i'}+A_iE_{i'j'}+\mu E_{ij}+A_{i'}E_{ij}+E_{ij}E_{i'j'}]$$
$$= \mu^2 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0$$
$$= \mu^2 \text{ for similar reasons as for the other expectations of the P's.}$$

Two approaches can be used to define the model in terms of expected values:

(1) in terms of P_{ij}'s and

(2) in terms of the model underlying the P_{ij}'s.

The second approach is usually used as in the example. In this case it is sometimes easier to write the model in terms of expected values of elements of the model. These can be found by implication from the model defined in terms of the P_{ij}'s.

For example in this model only the following definitions are needed:

$$E(P_{ij}) = \mu$$

$$E(P_{ij}^2) = \mu^2 + \sigma_A^2 + \sigma_E^2$$

$$E(P_{ij}P_{ij'}) = \mu^2 + \sigma_A^2 \text{ , and}$$

$$E(P_{ij}P_{i'j'}) = \mu^2 .$$

A second and more informative set of definitions is:

$$E(A_i) = 0 \text{ , } E(E_{ij}) = 0 \text{ ,}$$

$$E(A_i^2) = \sigma_A^2 \text{ , } E(E_{ij}^2) = \sigma_E^2 \text{ ,}$$

$$E(A_iA_{i'}) = 0 \text{ , } E(E_{ij}E_{ij'}) = 0, \text{ } E(E_{ij}E_{i'j'}) = 0 \text{ ,}$$

$E(A_iE_{ij}) = 0$, and $E(A_iE_{i'j}) = 0$ which are equivalent to the expectations in terms of the P_{ij}.

Fewer definitions are needed for the first method but the definitions are exactly equivalent.

A third way of completing the model which implies stating the properties of the elements of the equation of the model is to say that: 1) the A_i are independent ($\sigma_{A_iA_j} = 0$), identically distributed (all A_i from distribution with same variance and mean) with mean $\mu_A = 0$ and variance, σ_A^2; 2) the E_{ij} are independent, identically distributed

with mean, $\mu_E = 0$ and variance, σ^2_E; and 3) the A's and E's are independent. Thus the A's are IID(0, σ^2_A); the E's are IID(0, σ^2_E) and $E(A_i E_{jk}) = 0$ for all possible A's and E's. This method is the usual one for describing the model in research papers although both the first and second methods are more informative for working with expected values.

CHAPTER 37

REPEATABILITY

Repeatability is defined as the ratio of the variance due to animal effects to the total, or equivalently, phenotypic variance, that is:

$$r = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_E^2} \quad .$$

Repeatability also represents the fraction of the difference from the mean in one record which is expected in another record on the same animal. Repeatability thus is the regression coefficient for a subsequent record on a previous record and with equal variances for all records is also the correlation between records on the same animal.

The simple model is:

$P_{ij} = \mu + A_i + E_{ij}$, where

P_{ij} = j^{th} record on i^{th} animal,

A_i = effect on P_{ij} of the animal (genetic plus permanent environmental effects), and

E_{ij} = random temporary environmental effect

$i = 1, \ldots, B$ with B = number of animals

$j = 1, \ldots, n_i$ with n_i = number of records on i^{th} animal.

With method two of completing the model:

$$E(A_i) = \mu_A = 0 \qquad E(A_i^2) = \sigma_A^2 \qquad E(A_iA_j) = 0 \qquad E(A_iE_{ij}) = 0$$

$$E(A_iE_{i'j}) = 0 \qquad E(E_{ij}) = \mu_E = 0 \qquad E(E_{ij}^2) = \sigma_E^2 \qquad E(E_{ij}E_{ij'}) = 0$$

$$E(E_{ij}E_{i'j'}) = 0$$

These imply the following equivalent expressions for the first method of completing the model.

$$E(P_{ij}) = \mu_P = \mu \qquad E(P_{ij}^2) = \mu^2 + \sigma_A^2 + \sigma_E^2$$

$$E(P_{ij}P_{ij'}) = \mu^2 + \sigma_A^2 \qquad E(P_{ij}P_{i'j'}) = \mu^2$$

The third way of completing the model is to state that the A's are IID(0, σ_A^2), the E's are IID(0, σ_E^2) and the A's and E's are mutually uncorrelated.

ESTIMATION OF REPEATABILITY BY REGRESSION

The expectations of sums of squares and products used for estimating repeatability from regression and correlation coefficients are for $n_i = 2$ records for each of $B = n$ animals, $i = 1,..., n$ and $j = 1$ and 2:

$$E(\sum_i P_{i1}^2) = n\mu^2 + n\sigma_A^2 + n\sigma_E^2$$

$$E(P_{.1}^2/n) = n\mu^2 + \sigma_A^2 + \sigma_E^2$$

$$E(\Sigma P_{i2}^2) = n\mu^2 + n\sigma_A^2 + n\sigma_E^2$$

$$E(P_{.2}^2/n) = n\mu^2 + \sigma_A^2 + \sigma_E^2$$

$$E(\Sigma P_{i1}P_{i2}) = n\mu^2 + n\sigma_A^2$$

$$E(P_{.1}P_{.2}/n) = n\mu^2 + \sigma_A^2$$

Thus, where ˆ indicates an estimate, the estimates of variances and covariances are:

$$\hat{\sigma}^2_{P_1} = \left(\sum_{i=1}^{n} P_{i1}^2 - P_{.1}^2/n \right) / (n-1) \; ; \qquad \hat{\sigma}^2_{P_2} = \left(\sum_{i=1}^{n} P_{i2}^2 - P_{.2}^2/n \right) / (n-1)$$

$$\hat{\sigma}_{P_1 P_2} = \left(\sum_{i=1}^{n} P_{i1} P_{i2} - P_{.1} P_{.2}/n \right) / (n-1)$$

The expectations are: $E(\hat{\sigma}^2_{P_1}) = E(\hat{\sigma}^2_{P_2}) = \sigma^2_A + \sigma^2_E$ and $E(\hat{\sigma}_{P_1 P_2}) = \sigma^2_A$.

The estimate of repeatability by regression of second record on first record is:

$$r = \frac{\hat{\sigma}_{P_1 P_2}}{\hat{\sigma}^2_{P_1}} \qquad \text{where the expectation by parts is } \frac{\sigma^2_A}{\sigma^2_A + \sigma^2_E} \; .$$

The variance of the estimate is:

$$\sigma^2_r = \left(\hat{\sigma}^2_{P_2} / \hat{\sigma}^2_{P_1} - b^2 \right) / (n-2) \; .$$

The estimate of repeatability by correlation between first and second records is:

$$r = \frac{\hat{\sigma}_{P_1 P_2}}{\sqrt{\hat{\sigma}^2_{P_1} \, \hat{\sigma}^2_{P_2}}} \qquad \text{where the expectation taken separately for each part is:}$$

$$\frac{\sigma^2_A}{\sqrt{(\sigma^2_A + \sigma^2_E)(\sigma^2_A + \sigma^2_E)}} = \frac{\sigma^2_A}{\sigma^2_A + \sigma^2_E} \; .$$

Although the expectations by parts are the same for both the regression and correlation coefficients, the regression and correlation coefficients are rarely equal.

If some animals are not allowed to have a second record because of a low first record then $E(\hat{\sigma}^2_{P_1})$ from records of only animals that were allowed to have second records is less than $\sigma^2_A + \sigma^2_E$ by a factor, $k < 1$; i.e., $E(\hat{\sigma}^2_{P_1}) = k(\sigma^2_A + \sigma^2_E)$ where k depends on the intensity of selection. Fortunately $E(\hat{\sigma}_{P_1 P_2})$ is also less than σ^2_A by the same factor, k; i.e., $E(\hat{\sigma}_{P_1 P_2}) = k\,\sigma^2_A$. Thus the regression estimate is unbiased by selection on first records. The correlation estimate is biased since $E(\hat{\sigma}^2_{P_2})$ from records of only animals allowed second records is not $k(\sigma^2_A + \sigma^2_E)$ but is $k\sigma^2_A + \sigma^2_E$.

In the unusual situation when the pairs of first and second records depend on the size of the second record both regression and correlation estimates of repeatability will be biased.

Example of Computing Repeatability by Regression and Correlation.

The following set of first and second records on 10 animals was drawn from a population with $\mu = 500$, $\sigma_A = 20$, and $\sigma_E = 10$, i.e., $r = \dfrac{400}{400 + 100} = .80$.

Animal (i)	P_{i1}	P_{i2}
1	504	533
2	542	548
3	523	522
4	471	484
5	505	495
6	543	543
7	500	495
8	460	479
9	474	449
10	522	527
	$\overline{P_{.1}} = 5044$	$\overline{P_{.2}} = 5075$

$$\sum_{i=1}^{10} P_{i1}^2 = (504)^2 + (542)^2 + \cdots + (522)^2 = 2{,}551{,}784$$

$$\sum_{i=1}^{10} P_{i2}^2 = (533)^2 + (548)^2 + \cdots + (527)^2 = 2{,}584{,}803$$

$$\sum_{i=1}^{10} P_{i1}P_{i2} = (504)(533) + (542)(548) + \cdots + (522)(527) = 2{,}567{,}202$$

$$\hat{\sigma}^2_{P_1} = [2{,}551{,}784 - (5044)^2/10]/(10-1) = 843.6$$

$$\hat{\sigma}^2_{P_2} = [2{,}548{,}803 - (5075)^2/10]/(10-1) = 1026.8$$

$$\hat{\sigma}_{P_1P_2} = [2{,}567{,}202 - (5044)(5075)/10](10-1) = 819.2$$

By regression, $r = \dfrac{819.2}{843.6} = .97$ and by correlation, $r = \dfrac{819.2}{\sqrt{(843.6)(1026.8)}} = .88$

ESTIMATION OF REPEATABILITY FROM VARIANCE COMPONENTS

Now the i^{th} animal has n_i records, $j=1, \cdots, n_i$ and $n. = \Sigma n_i$. The expectations of the usual sums of squares used in estimating σ^2_A and σ^2_E are:

$$E\left(\sum_i \sum_j P_{ij}^2\right) = n.\mu^2 + n.\sigma^2_A + n.\sigma^2_E$$

$$E\left(\sum_i P_{i.}^2/n_i\right) = n.\mu^2 + n.\sigma^2_A + B\,\sigma^2_E$$

$$E\left(P_{..}^2/n.\right) = n.\mu^2 + \left(\frac{1}{n.}\right)\sum_i n_i^2\sigma^2_A + \sigma^2_E$$

The variance components to be estimated are replaced in the expectations by estimates of the variance components (i.e., ^'s are put on the symbols) and then equated to the computed sums of squares. The estimates are:

$$\hat{\sigma}^2_E = \left[\sum_i \sum_j P^2_{ij} - \sum_i P^2_{i.}/n_i\right] \; / \; (n. - B) \quad \text{where } E(\hat{\sigma}^2_E) = \sigma^2_E,$$

$$\hat{\sigma}^2_A = \left[\sum_i P^2_{i.}/n_i - P^2_{..}/n. - (B - 1)\,\hat{\sigma}^2_E\right] / \left[n. - (1/n.) \sum_i n_i^2\right] \quad \text{where } E(\hat{\sigma}^2_A) = \sigma^2_A \; .$$

Thus:

$$r = \frac{\hat{\sigma}^2_A}{\hat{\sigma}^2_A + \hat{\sigma}^2_E} \; .$$

The approximate variance of this estimate is:

$$\sigma^2_r \; - \; \frac{2(n.-1)\,(1-r)^2\,[1 + (k-1)r]^2}{k^2\,(n.-B)(B-1)}$$

where $k = [1/(B-1)]\,[n. - (1/n.)\,\Sigma\, n_i^2]$.

Example of Estimating Repeatability From Variance Components.

The following set of 10 records on 5 animals was drawn from a population with $\mu = 1000$, $\sigma_A = 10$, and $\sigma_E = 10$.

	Record (j)				
Animal (i)	1	2	3	$P_{i.}$	n_i
1	979	976	984	2939	3
2	988	1007	----	1995	2
3	994	----	----	994	1
4	1004	1017	----	2021	2
5	1015	1022	----	2037	2
				$P_{..} = 9986$	$n_. = 10$

$$\sum_{i=1}^{5}\sum_{j=1}^{n_i} P_{ij}^2 = (979)^2 + (976)^2 + (984)^2 + (988)^2 + \cdots + (1022)^2 = 9{,}974{,}516$$

$$\sum_{i=1}^{5} \frac{P_{i.}^2}{n_i} = \frac{(2939)^2}{3} + \frac{(1995)^2}{2} + \frac{(994)^2}{1} + \frac{(2021)^2}{2} + \frac{(2037)^2}{2} = 9{,}974{,}193$$

$$\frac{P_{..}^2}{n.} = (9986)^2/10 = 9{,}972{,}019$$

$$\sum_{i=1}^{5} n_i^2 = (3)^2 + (2)^2 + (1)^2 + (2)^2 + (2)^2 = 22 \; ; \quad \Sigma\, n_i^2/n. = 22/10 = 2.2$$

$$\hat{\sigma}_E^2 = (9{,}974{,}516 - 9{,}974{,}193)/(10-5) = 64.6$$

$$\hat{\sigma}_A^2 = [9{,}974{,}193 - 9{,}972{,}019 - (5-1)(64.6)]/(10-2.2) = 245.6$$

$$r = \frac{245.6}{245.6 + 64.6} = .79$$

CHAPTER 38

HERITABILITY

Heritability is defined as the ratio of the variance due to genetic effects to the phenotypic variance. That is, $h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_e^2)$. Most methods of estimating h^2 make the assumption that σ_g^2 is the variance of additive genetic effects. Biased estimates may result if the assumption is not true.

ESTIMATION OF HERITABILITY BY REGRESSION AND CORRELATION

Estimation of heritability by regression uses **n** pairs of records of relatives X and Y which have additive relationship, a_{xy}.

The model for the records is:

$$P_{xi} = \mu + g_{xi} + e_{xi} \quad \text{and} \quad P_{yi} = \mu + g_{yi} + e_{yi} \qquad (i = 1, ..., n)$$

$E(P_{xi}) = E(P_{yi}) = \mu$;

$E\left(P_{xi}^2\right) = E\left(P_{yi}^2\right) = \mu^2 + \sigma_g^2 + \sigma_e^2$ which imply $E(g_{xi}e_{xi}) = E(g_{yi}e_{yi}) = 0$;

$E(P_{xi}P_{yi'}) = \mu^2$, $E(P_{xi}P_{xi'}) = \mu^2$, $E(P_{yi}P_{yi'}) = \mu^2$ for $i \neq i'$;

$E(P_{xi}P_{yi}) = \mu^2 + a_{xy}\sigma_g^2$ which implies $E(g_{xi}e_{yi}) = E(g_{yi}e_{xi}) = E(e_{xi}e_{yi}) = 0$.

If other than additive gene effects contribute to genetic variance $E(P_{xi}P_{yi}) = \sum_k \sum_j a_{xy}^k d_{xy}^j \sigma_{kj}^2 + \mu^2$ where $j+k \geq 1$; d_{xy} is dominance relationship between relatives X and Y and σ_{kj}^2 is the variance of specific genetic effects due to the action of single genes at k loci and genotypes at j other loci as in Chapter 6.

The expectations of the sums of squares and crossproducts used to estimate h^2 from regression and correlation coefficients are:

$$E\left(\sum_i P_{xi}^2\right) = n\mu^2 + n\sigma_g^2 + n\sigma_e^2; \qquad E\left(P_{x.}^2/n\right) = n\mu^2 + \sigma_g^2 + \sigma_e^2;$$

$$E\left(\sum_i P_{yi}^2\right) = n\mu^2 + n\sigma_g^2 + n\sigma_e^2; \qquad E\left(P_{y.}^2/n\right) = n\mu^2 + \sigma_g^2 + \sigma_e^2;$$

$$E\left(\sum_i P_{xi}P_{yi}\right) = n\mu^2 + na_{xy}\sigma_g^2; \qquad E\left(\frac{P_{x.}P_{y.}}{n}\right) = n\mu^2 + a_{xy}\sigma_g^2 \ .$$

Thus:

$$\hat{\sigma}_x^2 = \left(\sum_i P_{x_i}^2 - \frac{P_{x.}^2}{n}\right) / (n-1) \text{ where } E(\hat{\sigma}_x^2) = \sigma_g^2 + \sigma_e^2 \ ;$$

$$\hat{\sigma}_y^2 = \left(\sum_i P_{y_i}^2 - \frac{P_{y.}^2}{n}\right) / (n-1) \text{ where } E(\hat{\sigma}_y^2) = \sigma_g^2 + \sigma_e^2 \ ; \text{ and}$$

$$\hat{\sigma}_{xy} = \left(\sum_i P_{x_i}P_{y_i} - \frac{P_{x.}P_{y.}}{n}\right) / (n-1) \text{ where } E(\hat{\sigma}_{xy}) = a_{xy}\,\sigma_g^2 \text{ if } \sigma_g^2 = \sigma_{10}^2 \ .$$

Remember that in general $E(\hat{\sigma}_{xy}) = a_{xy}\,\sigma_{10}^2 + d_{xy}\,\sigma_{01}^2 + a_{xy}^2\,\sigma_{20}^2 + \cdots$.

Then the regression of record of Y on record of X is: $b_{y \cdot x} = \hat{\sigma}_{xy}/\hat{\sigma}_x^2$.

The correlation of record of Y and record of X is: $r_{yx} = \dfrac{\hat{\sigma}_{xy}}{\sqrt{\hat{\sigma}_x^2 \, \hat{\sigma}_y^2}}$.

The estimate of h^2 from regression of the record of relative Y on relative X is:

$h^2 = \dfrac{1}{a_{xy}} b_{y \cdot x}$ where the expectation by parts (expectation of numerator and denominator parts is done separately) gives:

$$\left[\frac{1}{a_{xy}}\right]\left[\frac{a_{xy}\,\sigma_{10}^2}{\sigma_g^2 + \sigma_e^2}\right] = \frac{\sigma_{10}^2}{\sigma_g^2 + \sigma_e^2} \quad \text{for } \sigma_g^2 = \sigma_{10}^2 \text{ but otherwise gives}$$

$$\left[\frac{1}{a_{xy}}\right]\left[\frac{a_{xy}\,\sigma_{10}^2 + d_{xy}\,\sigma_{01}^2 + a_{xy}^2\,\sigma_{20}^2 + \cdots}{\sigma_g^2 + \sigma_e^2}\right] \text{ so that the bias will be :}$$

$$\frac{d_{xy}\,\sigma_{01}^2 + a_{xy}^2\,\sigma_{20}^2 + \cdots}{a_{xy}\,(\sigma_g^2 + \sigma_e^2)} \quad .$$

The variance of the regression estimate is (the square root is the standard error):

$$V(h^2) = \frac{1}{a_{xy}^2} V(b) = \frac{\hat{\sigma}_y^2 / \hat{\sigma}_x^2 - b^2}{a_{xy}^2 (n-2)}$$

From correlation the estimate of h^2 is:

$$h^2 = \frac{1}{a_{xy}} r_{yx}$$

where the expectation by parts is the same as by regression.

As in the case of estimating repeatability by regression the estimate of heritability is unbiased by selection on relative X as to whether a record on relative Y will be available, e.g., by selection of potential parents, X, to have progeny, Y.

Example Of Estimating Heritability By Regression And Correlation

The following set of 20 paired records on progeny and parents was drawn from a population with $\mu = 150$, $\sigma_g = 10$, and $\sigma_e = 20$.

Pair (i)	Parent (X)	Progeny (Y)	Pair (i)	Parent (X)	Progeny (Y)
1	194	172	11	168	148
2	164	175	12	164	173
3	147	142	13	202	169
4	165	159	14	129	145
5	181	148	15	194	184
6	155	164	16	148	176
7	142	157	17	171	169
8	153	166	18	148	105
9	159	144	19	167	162
10	138	113	20	146	169
				$\overline{P_{x.}} = 3235$	$\overline{P_{y.}} = 3140$

$$\sum_i P_{xi}^2 = (194)^2 + (164)^2 + \cdots + (146)^2 = 530{,}505$$

$$\sum_i P_{yi}^2 = (172)^2 + (175)^2 + \cdots + (169)^2 = 500{,}806$$

$$\sum_i P_{xi}P_{yi} = (194)(172) + (164)(175) + \cdots + (146)(169) = 511{,}557$$

$$\hat{\sigma}_x^2 = [530{,}505 - (3235)^2/20]/19 = 381.3$$

$$\hat{\sigma}_y^2 = [500{,}806 - (3140)^2/20]/19 = 411.9$$

$$\hat{\sigma}_{xy} = [511{,}557 - (3235)(3140)/20]/19 = 192.7$$

Since $a_{xy} = 1/2$, by regression, $h^2 = \frac{2(192.7)}{381.3} = 1.01$ with standard error .43 and by correlation, $h^2 = \frac{2(192.7)}{\sqrt{(381.3)(411.9)}} = .97$.

This example illustrates the fact that although the upper and lower limits of h^2 are 1 and 0, estimates from regression and correlation can be outside the limits.

ESTIMATION OF HERITABILITY BY VARIANCE COMPONENTS

One-Way Classification

The one-way classification model is:

$$P_{ij} = \mu + b_i + w_{ij}, \text{ where}$$

μ is a constant,

b_i is an effect in common to members of the i^{th} genetic group,

w_{ij} is a random effect associated with the j^{th} member of the i^{th} group.

$i=1, ..., B$; $j=1, ..., n_i$.

The b's are $IID\left(0,\sigma_b^2\right)$, the w's are $IID\left(0,\sigma_w^2\right)$ and the b's and w's are mutually uncorrelated. The variance among groups, $\sigma_b^2 = a_{ii}\sigma_{10}^2 + d_{ii}\sigma_{01}^2 + \cdots$, can be shown to be the covariance between animals in the same group with a_{ii}, the additive relationship between pairs of the i^{th} group, and d_{ii}, the corresponding dominance relationship.

From genetic theory $a_{ii}\sigma_{10}^2 + d_{ii}\sigma_{01}^2 + \cdots = Cov\left(P_{ij},P_{ij'}\right)$ which by definition equals $E[(P_{ij} - \mu)(P_{ij'} - \mu)] = E(P_{ij},P_{ij'}) - \mu^2$. Thus $\sigma_b^2 + \mu^2 = E(P_{ij}P_{ij'})$. Note that $d_{ii'} = a_{ii'} = 0$ for $i \neq i'$, i.e., between groups that are not related to each other. Because $\sigma_P^2 = \sigma_b^2 + \sigma_w^2$ and also $\sigma_P^2 = \sigma_g^2 + \sigma_e^2$; then $\sigma_w^2 = \sigma_g^2 + \sigma_e^2 - \sigma_b^2$.

If $\sigma_g^2 = \sigma_{10}^2$, then,

$$\sigma_w^2 = (1 - a_{ii})\sigma_{10}^2 + \sigma_e^2 \quad \text{and} \quad \sigma_b^2 = a_{ii}\sigma_{10}^2 \ .$$

In terms of the P's the expected values are:

$$E(P_{ij}) = \mu \ ; \quad E\left(P_{ij}^2\right) = \mu^2 + \sigma_b^2 + \sigma_w^2$$

$E(P_{ij}P_{ij'}) = \mu^2 + \sigma_b^2$; $E(P_{i'j'}P_{ij}) = \mu^2$, $E(P_{ij}P_{i'j}) = \mu^2$. These definitions of the model lead to the following expectations of the three sums of squares (also called quadratics) usually used to estimate σ_b^2 and σ_w^2:

$$E\left(\sum_i \sum_j P_{ij}^2\right) = n.\mu^2 + n.\sigma_b^2 + n.\sigma_w^2$$

$$E\left(\sum_i \frac{P_{i.}^2}{n_i}\right) = n.\mu^2 + n.\sigma_b^2 + B\,\sigma_w^2$$

$$E\left(\frac{P_{..}^2}{n.}\right) = n.\mu^2 + \frac{1}{n.}\left(\sum_i n_i^2\right)\sigma_b^2 + \sigma_w^2 \ .$$

The estimates are obtained by equating the quadratics to their expected values:

$$\hat{\sigma}_w^2 = \left(\sum_i \sum_j P_{ij}^2 - \sum_i \frac{P_{i.}^2}{n_i}\right) / (n. - B)$$

$$\hat{\sigma}_b^2 = \left(\sum_i \frac{P_{i.}^2}{n_i} - \frac{P_{..}^2}{n.} - (B-1)\,\hat{\sigma}_w^2\right) / \left(n. - \frac{1}{n.}\sum_i n_i^2\right)$$

The estimate of the intraclass correlation is:

$$t = \hat{\sigma}_b^2 \Big/ \left(\hat{\sigma}_b^2 + \hat{\sigma}_w^2\right) \text{ and the estimate of heritability is: } h^2 = \left(1/a_{ii}\right)t \ .$$

The approximate variance of this estimate is:

$$V(h^2) = V(t)/a_{ii}^2 \text{ where}$$

$$V(t) = \frac{2(n.-1)\ (1-t)^2\ [1 + (k-1)\ t]^2}{k^2\ (n. - B)(B - 1)} \text{ with } k = \left[n. - \left(\Sigma n_i^2\right) / n.\right] / [B - 1] \ .$$

Note that heritability is a simple multiple of the intra-class correlation coefficient, i.e., $h^2 = \frac{t}{a_{ii}}$ which is the reason that this method is often called the intra-class correlation method of estimating heritability.

The Intra-Class Correlation Coefficient

The term, intra-class correlation coefficient, is, on examining all the adjectives, simply the correlation between records of any pair of animals in the same class--in this case in the same genetic group, i.e., $t = \frac{Cov(P_{ij},P_{ij'})}{\sqrt{Var(P_{ij})\ Var(P_{ij'})}}$.

But $Cov(P_{ij},P_{ij'}) = \sigma_b^2$ and $Var(P_{ij}) = Var(P_{ij'}) = \sigma_b^2 + \sigma_w^2$. Thus $t = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}$.

The genetic groups will often be groups with the same sire (paternal sib groups, $a_{ii} = 1/4$, $d_{ii} = 0$), with the same sire and dam (full sib groups, $a_{ii} = 1/2$, $d_{ii} = 1/4$), or with the same dam (maternal sib groups, $a_{ii} = 1/4$, $d_{ii} = 0$). For this model to be correct, the groups cannot be related to each other, e.g., no sire can be the sire of more than one full sib group.

Example of Estimating Heritability for the One-way Classification Model

The following set of 75 records from 25 paternal sib groups was drawn from a population with $\mu = 150$, $\sigma_b^2 = 50$, and $\sigma_w^2 = 450$.

Sire	Progeny						Sire	Progeny					
(i)	1	2	3	4	5	$P_{i.}$	(i)	1	2	3	4	5	$p_{i.}$
1	132	138	----	----	----	270	14	156	178	165	153	145	797
2	115	135	156	----	----	406	15	125	172	142	117	----	556
3	181	----	----	----	----	181	16	157	152	----	----	----	309
4	146	133	165	144	163	751	17	152	142	161	----	----	455
5	143	148	173	147	----	611	18	153	----	----	----	----	153
6	128	113	----	----	----	241	19	152	140	170	165	144	771
7	159	157	138	----	----	454	20	135	169	148	137	----	589
8	175	----	----	----	----	175	21	156	151	----	----	----	307
9	126	160	162	130	172	750	22	165	162	111	----	----	438
10	170	129	134	165	----	598	23	160	----	----	----	----	160
11	140	164	----	----	----	304	24	132	119	153	129	117	650
12	138	168	128	----	----	434	25	138	150	144	152	----	584
13	154	----	----	----	----	154							

$$\sum_i \sum_j P_{ij}^2 = 132^2 + 138^2 + 115^2 + \cdots + 152^2 = 1{,}663{,}334$$

$$\sum_i \frac{P_{i.}^2}{n_i} = \frac{(270)^2}{2} + \frac{(406)^2}{3} + \frac{(181)^2}{1} + \cdots + \frac{(584)^2}{4} = 1{,}649{,}806.5$$

$$\frac{P_{..}^2}{n.} = \frac{(11098)^2}{75} = 1{,}642{,}208.1, \quad \sum n_i^2 = 2^2 + 3^2 + 1^2 + \cdots + 4^2 = 275$$

$$\hat{\sigma}_w^2 = (1{,}663{,}334 - 1{,}649{,}806.5) / (75 - 25) = 270.5$$

$$\hat{\sigma}_b^2 = [1{,}649{,}806.5 - 1{,}642{,}208.1 - (25-1)(270.5)] / [75 - 275/75] = 15.5$$

Then,

$$h^2 = \frac{4(15.5)}{15.5 + 270.5} = .22 \ .$$

Two-Way Nested Classification

Estimation of Heritability

This model is appropriate when sires are mated to many dams, but each dam is mated to only one sire with one or more progeny per dam:

$$P_{ijk} = \mu + s_i + d_{ij} + w_{ijk}, \text{ where}$$

μ is a constant,

s_i is the effect common to all animals with the $i^{\underline{th}}$ sire,

d_{ij} is the additional effect common to all animals with the $j^{\underline{th}}$ dam mated to the $i^{\underline{th}}$ sire, i.e., $s_i + d_{ij} = ij^{\underline{th}}$ full sib effect so that $d_{ij} = (\text{full sib effect})_{ij} - s_i$,

w_{ijk} is a random effect associated with the record of the $k^{\underline{th}}$ member of the $ij^{\underline{th}}$ full sib group.

$i=1, ..., S$; $j=1, ..., m_i$; $k=1, ..., n_{ij}$; $m.=D$, the total number of dams, and with the s's, $\text{IID}\left(0,\sigma_s^2\right)$; the d's, $\text{IID}\left(0,\sigma_d^2\right)$; the w's, $\text{IID}\left(0,\sigma_w^2\right)$ and the s's, d's, and w's mutually uncorrelated.

Note that:

$\sigma_s^2 = \frac{1}{4}\sigma_{10}^2 + \frac{1}{16}\sigma_{20}^2 + \cdots$ is the covariance among paternal half sibs;

$\sigma_d^2 = \sigma_{FS}^2 - \sigma_s^2 = \frac{1}{4}\sigma_{10}^2 + \frac{1}{16}\sigma_{20}^2 + \cdots + \frac{1}{4}\sigma_{01}^2 + \cdots$ is the covariance among full sibs, σ_{FS}^2, minus the covariance among paternal half sibs, σ_s^2 (if maternal effects exist, then σ_d^2 also includes σ_m^2, the variance of maternal effects) and

$\sigma_w^2 = \sigma_g^2 + \sigma_e^2 - \sigma_s^2 - \sigma_d^2$ is total phenotypic variance, $\sigma_g^2 + \sigma_e^2$, minus σ_s^2 and σ_d^2.

If $\sigma_g^2 = \sigma_{10}^2 + \sigma_{01}^2$ then $\sigma_w^2 = \frac{1}{2}\sigma_{10}^2 + \frac{3}{4}\sigma_{01}^2 + \sigma_e^2$.

In terms of different combinations of the P's the expected values are:

$$E(P_{ijk}) = \mu$$

$$E(P_{ijk}^2) = \mu^2 + \sigma_s^2 + \sigma_d^2 + \sigma_w^2$$

$$E(P_{ijk}P_{ijk'}) = \mu^2 + \sigma_s^2 + \sigma_d^2 \qquad \text{(records of full sibs)}$$

$$E(P_{ijk}P_{ij'k'}) = \mu^2 + \sigma_s^2 \qquad \text{(records of paternal half sibs)}$$

$$E(P_{ijk}P_{i'j'k'}) = \mu^2 \qquad \text{(records of unrelated animals).}$$

These definitions of the model lead to the expectations of the four quadratics used to estimate σ_s^2, σ_d^2, and σ_w^2 :

$$E\left(\sum_i \sum_j \sum_k P_{ijk}^2\right) = n..\mu^2 + n..\sigma_s^2 + n..\sigma_d^2 + n..\sigma_w^2$$

$$E\left(\sum_i \sum_j \frac{P_{ij.}^2}{n_{ij}}\right) = n..\mu^2 + n..\sigma_s^2 + n..\sigma_d^2 + D\,\sigma_w^2$$

$$E\left(\sum_i \frac{P_{i..}^2}{n_{i.}}\right) = n..\mu^2 + n..\sigma_s^2 + \sum_i \frac{\sum_j n_{ij}^2}{n_{i.}}\sigma_d^2 + S\,\sigma_w^2$$

$$E\left(\frac{P_{...}^2}{n..}\right) = n..\mu^2 + \frac{1}{n..}\sum_i n_{i.}^2\,\sigma_s^2 + \frac{1}{n..}\sum_i \sum_j n_{ij}^2\,\sigma_d^2 + \sigma_w^2$$

When the quadratics are equated to their expected values the estimates are:

$$\hat{\sigma}^2_w = \left(\sum_i \sum_j \sum_k P^2_{ijk} - \sum_i \sum_j \frac{P^2_{ij.}}{n_{ij}} \right) \Big/ (n.. - D)$$

$$\hat{\sigma}^2_d = \left(\sum_i \sum_j \frac{P^2_{ij.}}{n_{ij}} - \sum_i \frac{P^2_{i..}}{n_{i.}} - (D-S)\, \hat{\sigma}^2_w \right) \Big/ \left(n.. - \sum_i \frac{\sum_j n^2_{ij}}{n_{i.}} \right)$$

$$\hat{\sigma}^2_s = \frac{\sum_i \frac{P^2_{i..}}{n_{i.}} - \frac{P^2_{...}}{n_{..}} - \left(\sum_i \frac{\sum_j n^2_{ij}}{n_{i.}} - \frac{1}{n_{..}} \sum_i \sum_j n^2_{ij} \right) \hat{\sigma}^2_d - (S-1)\, \hat{\sigma}^2_w}{\left(n_{..} - \frac{1}{n_{..}} \sum_i n^2_{i.} \right)}$$

Then $\hat{\sigma}^2_P = \hat{\sigma}^2_s + \hat{\sigma}^2_d + \hat{\sigma}^2_w$.

If $\sigma^2_{20} = \sigma^2_{30} = \cdots = 0$, the estimate of heritability from the sire variance is:

$$h^2_s = \frac{4\hat{\sigma}^2_s}{\hat{\sigma}^2_p} \quad \text{i.e., } \hat{\sigma}^2_{10} = 4\, \hat{\sigma}^2_s \ .$$

If σ^2_{01} can be assumed to equal zero then two other estimates that use the dam component of variance are:

$$h^2_d = \frac{4\hat{\sigma}^2_d}{\hat{\sigma}^2_P} \text{ and } h^2_{s+d} = \frac{2\left(\hat{\sigma}^2_s + \hat{\sigma}^2_d\right)}{\hat{\sigma}^2_P} \ .$$

If σ^2_{20} and other second and higher order components, σ^2_{11} etc., are zero then:

$$\hat{\sigma}^2_{01} = 4\left(\hat{\sigma}^2_d - \hat{\sigma}^2_s\right) \ .$$

Example of Estimating Heritability by Variance Components for the Two-way Nested Classification Model: Progeny with One Record Each, Nested in Dams, Nested in Sires

The following set of 50 records from 10 dams mated to 5 sires was drawn from a population with $\mu = 150$, $\sigma_s^2 = 50$, $\sigma_d^2 = 75$ and $\sigma_w^2 = 375$.

Sire (i)	Dam (j)	Prog (k)	P_{ijk}	Sire (i)	Dam (j)	Prog (k)	P_{ij}	Sire (i)	Dam (j)	Prog (k)	P_{ijk}
1	1	1	137	3	1	1	169	5	1	1	138
1	1	2	166	3	1	2	163	5	1	2	150
		$P_{11.}$ =	303			$P_{31.}$ =	332			$P_{51.}$ =	288
1	2	1	142	3	2	1	126	5	2	1	136
1	2	2	103	3	2	2	173	5	2	2	142
1	2	3	125	3	2	3	176	5	2	3	128
1	2	4	153	3	2	4	154	5	2	4	145
1	2	5	180	3	2	5	169	5	2	5	168
1	2	6	170	3	2	6	179	5	2	6	149
1	2	7	157	3	2	7	178	5	2	7	152
1	2	8	153	3	2	8	191	5	2	8	137
		$P_{12.}$ =	1183			$P_{32.}$ =	1346			$P_{52.}$ =	1157
		$P_{1..}$ =	1486			$P_{3..}$ =	1678			$P_{5..}$ =	1445
2	1	1	110	4	1	1	144				
2	1	2	112	4	1	2	144				
		$P_{21.}$ =	222			$P_{41.}$ =	288				
2	2	1	190	4	2	1	170				
2	2	2	199	4	2	2	147				
2	2	3	191	4	2	3	147				
2	2	4	171	4	2	4	161				
2	2	5	156	4	2	5	159				
2	2	6	158	4	2	6	125				
2	2	7	171	4	2	7	128				
2	2	8	184	4	2	8	128				
		$P_{22.}$ =	1420			$P_{42.}$ =	1165				
		$P_{2..}$ =	1642			$P_{4..}$ =	1453			$P_{...}$ =	7704

$$\sum_i \sum_j \sum_k P_{ijk}^2 = 137^2 + 166^2 + 142^2 + \cdots + 153^2 + 110^2 + \cdots + 137^2 = 1{,}211{,}444$$

$$\sum_i \sum_j \frac{P_{ij.}^2}{n_{ij}} = \frac{303^2}{2} + \frac{1183^2}{8} + \frac{222^2}{2} + \cdots + \frac{1157^2}{8} = 1{,}199{,}037.4$$

$$\sum_i \frac{P_{i..}^2}{n_{i.}} = \frac{1486^2}{10} + \frac{1642^2}{10} + \cdots + \frac{1445^2}{10} = 1{,}191{,}927.8$$

$$\frac{P_{...}^2}{n..} = \frac{7704^2}{50} = 1{,}187{,}032.3$$

Note $n.. = 50$, $D = 10$, $S = 5$.

$$\sum_i \frac{\sum_j n_{ij}^2}{n_{i.}} = \frac{2^2 + 8^2}{10} + \frac{2^2 + 8^2}{10} + \cdots + \frac{2^2 + 8^2}{10} = 34$$

$$\sum_i \sum_j n_{ij}^2 = 2^2 + 8^2 + 2^2 + \cdots + 8^2 = 340,$$

$$\sum n_{i.}^2 = 10^2 + \cdots + 10^2 = 500$$

$$\hat{\sigma}_w^2 = (1{,}211{,}444 - 1{,}199{,}037.4)/(50-10) = 310.2$$

$$\hat{\sigma}_d^2 = [1{,}199{,}037.4 - 1{,}191{,}927.8 - (10-5)(310.2)]/(50-34) = 347.4$$

$$\hat{\sigma}_s^2 = \frac{[1{,}191{,}927.8 - 1{,}187{,}032.3 - (34 - 340/50)(347.4) - (5-1)(310.2)]}{[50 - 500/50]} = -144.9$$

$$\hat{\sigma}_P^2 = -144.9 + 347.4 + 310.2 = 512.7$$

$$h_s^2 = 4(-144.9)/512.7 = -1.13$$

This estimate illustrates the point that estimates of h^2 by this method may be outside of the theoretical limits of 0 and 1.

In this case, $\hat{\sigma}^2_{10} = 4(-144.9) = -579.6$ and $\hat{\sigma}^2_{01} = 4[347.4 - (-144.9)] = 1969.2$, which is also obviously unreasonable because that is greater than the total phenotypic variance.

Two-Way Nested Classification

Joint Estimation of Heritability and Repeatability

The model applies when a sire has many progeny each with one or more records and dams are assumed to be unrelated:

$$P_{ijk} = \mu + s_i + c_{ij} + w_{ijk}, \text{ where}$$

μ is a constant,

s_i is the effect common to all animals with the $i^{\underline{th}}$ sire,

c_{ij} is the additional effect common to all records of the $j^{\underline{th}}$ progeny sired by the $i^{\underline{th}}$ sire, i.e., $s_i + c_{ij}$ corresponds to animal effect of repeatability model,

w_{ijk} is what is left over, a random effect associated with the $k^{\underline{th}}$ record of the $j^{\underline{th}}$ progeny of the $i^{\underline{th}}$ sire.

$i = 1, ..., S$; $j = 1, ..., m_i$; $m. = C$, the number of animals; and $k = 1, ..., n_{ij}$.

Note that this is the same statistical model as the previous one except that c_{ij} appears rather than d_{ij} and C rather than D. The assumptions about the elements of the model are the same except that now σ^2_c is interpreted differently from σ^2_d.

Again,

$$\sigma^2_s = \frac{1}{4}\sigma^2_{10} + \frac{1}{16}\sigma^2_{20} + \cdots$$

But now,

$\sigma^2_c = \sigma^2_A - \sigma^2_s$ with $\sigma^2_A = \sigma^2_g + \sigma^2_{PE}$ where σ^2_{PE} is the variance of permanent environmental effects. Thus, $\sigma^2_A = \sigma^2_s + \sigma^2_c$.

Now, $\sigma_w^2 = \sigma_{TE}^2$ is the variance of temporary environmental effects where in the previous heritability models, $\sigma_w^2 = \sigma_{TE}^2 + \sigma_{PE}^2 +$ some genetic variance.

If $\sigma_g^2 = \sigma_{10}^2$, then $\sigma_c^2 = \frac{3}{4}\sigma_{10}^2 + \sigma_{PE}^2$.

The expectations of different combinations of the P's are:

$E(P_{ijk}) = \mu$

$E(P_{ijk}^2) = \mu^2 + \sigma_s^2 + \sigma_c^2 + \sigma_w^2$

$E(P_{ijk}P_{ijk'}) = \mu^2 + \sigma_s^2 + \sigma_c^2$ (records on the same animal)

$E(P_{ijk}P_{ij'k'}) = \mu^2 + \sigma_s^2$ (records of paternal half sibs)

$E(P_{ijk}P_{i'j'k'}) = \mu^2$ (records of unrelated animals).

The same four quadratics as before are used to estimate σ_s^2, σ_c^2, and σ_w^2 . Their expectations lead to the estimates:

$$\hat{\sigma}_w^2 = \left(\sum_i \sum_j \sum_k P_{ijk}^2 - \sum_i \sum_j \frac{P_{ij.}^2}{n_{ij}} \right) / (n.. - C)$$

$$\hat{\sigma}_c^2 = \left(\sum_i \sum_j \frac{P_{ij.}^2}{n_{ij}} - \sum_i \frac{P_{i..}^2}{n_{i.}} - (C - S)\,\hat{\sigma}_w^2 \right) / \left(n.. - \sum_i \frac{\sum_j n_{ij}^2}{n_{i.}} \right)$$

$$\hat{\sigma}_s^2 = \frac{\left[\sum_i \frac{P_{i..}^2}{n_{i.}} - \frac{P_{...}^2}{n..} - \left(\sum_i \frac{\sum_j n_{ij}^2}{n_{i.}} - \frac{1}{n..} \sum_i \sum_j n_{ij}^2 \right) \hat{\sigma}_c^2 - (S-1)\,\hat{\sigma}_w^2 \right]}{\left(n.. - \frac{1}{n..} \sum_i n_{i.}^2 \right)}$$

The estimate of phenotypic variance is:

$$\hat{\sigma}_P^2 = \hat{\sigma}_w^2 + \hat{\sigma}_c^2 + \hat{\sigma}_s^2 .$$

The estimate of heritability is:

$$h^2 = \frac{4\,\hat{\sigma}_s^2}{\hat{\sigma}_P^2} .$$

The estimate of repeatability is:

$$r = \frac{\hat{\sigma}_c^2 + \hat{\sigma}_s^2}{\hat{\sigma}_P^2} .$$

The within sire estimate of repeatability is:

$$r_{\text{within sire}} = \frac{\hat{\sigma}_c^2}{\hat{\sigma}_c^2 + \hat{\sigma}_w^2} .$$

Because the computing procedure is the same as for the previous example no example is given. Instead, as an exercise in simulating models and in computing estimates for the two-way nested model for joint estimation of h^2 and r the following problem is given.

Simulate the Repeatability and Heritability Model; Two-way Nested Classification

$$P_{ijk} = \mu + s_i + c_{ij} + w_{ijk}, \text{ where}$$

μ is a constant,

s_i is a random effect common to all animals with the i^{th} sire,

c_{ij} is a random effect common to all records of the j^{th} animal sired by the i^{th} sire, and

w_{ijk} is a random effect common to the k^{th} record of the ij^{th} animal.

Assume:

$$\sigma_s^2 = (1/4)\ \sigma_{10}^2 = (1/4)\ h^2\sigma_P^2$$

$$\sigma_c^2 = (r - h^2/4)\sigma_P^2, \text{ where } \sigma_s^2 + \sigma_c^2 = \sigma_A^2 \text{ , and}$$

$$\sigma_w^2 = \sigma_e^2 \ .$$

Note that $\sigma_s^2 + \sigma_c^2 = r\sigma_P^2$ and $r = \left(\sigma_s^2 + \sigma_c^2\right) / \sigma_P^2$.

For the simulation problem let:

$\mu = 500,$

$\sigma_s^2 = 400,$

$\sigma_c^2 = 2500,$ and

$\sigma_w^2 = 1600.$

Generate a sample of records for this model (see, Chapters 40 and 41) using random numbers and random normal deviates ($\mu = 0, \sigma = 1$) according to the pattern indicated on the following page. From these records estimate σ_s^2, σ_c^2, σ_w^2, h^2 and r .

Table for simulating two-way nested classification:

			Random Numbers			Random Values (nearest whole number)							
i	j	k	S_i	C_{ij}	W_{ijk}	μ	+	s_i	+	c_{ij}	+	w_{ijk}	= P_{ijk}
1	1	1											
1	1	2	-	-									
1	2	1	-										
2	1	1											
2	2	1	-										
2	2	2	-	-									
2	2	3	-	-									
2	3	1	-										
2	3	2	-	-									
3	1	1											
3	1	2	-	-									
3	2	1	-										
3	2	2	-	-									
3	3	1	-										
3	4	1	-										
4	1	1											
4	1	2	-	-									
4	2	1	-										
4	3	1	-										
4	3	2	-	-									

Two-Way Cross Classification Model

The mating pattern for this model is that sires are mated to many dams, dams are mated to many sires, and each progeny has only one record.

$$P_{ijk} = \mu + s_i + d_j + (sd)_{ij} + w_{ijk}, \text{ where}$$

μ is a constant,

s_i is the effect common to all animals with $i^{\underline{th}}$ sire,

d_j is the effect common to all animals with $j^{\underline{th}}$ dam,

$(sd)_{ij}$ is the effect (difference from $s_i + d_j$) common to all animals in the $ij^{\underline{th}}$ full sib group, so that $(sd)_{ij} = (\text{full sib effect})_{ij} - s_i - d_j$, and

w_{ijk} is a random effect associated with the record of the $k^{\underline{th}}$ member of the $ij^{\underline{th}}$ full sib group.

$i = 1, ..., S$; $j = 1, ..., D$; $k = 1, ..., n_{ij}$; and C = no. of matings or n_{ij}'s > 0.

The s's are $IID(0,\sigma_s^2)$, the d's are $IID(0,\sigma_d^2)$, the (sd)'s are $IID(0,\sigma_{(sd)}^2)$, the w's are $IID(0,\sigma_w^2)$ and the s's, d's, (sd)'s and w's are mutually uncorrelated.

$\sigma_s^2 = \frac{1}{4}\sigma_{10}^2 + \frac{1}{16}\sigma_{20}^2 + \cdots$, is the covariance among paternal half sibs;

$\sigma_d^2 = \frac{1}{4}\sigma_{10}^2 + \frac{1}{16}\sigma_{20}^2 + \cdots$, is the covariance among maternal half sibs;

$\sigma_{(sd)}^2 = \sigma_{FS}^2 - \sigma_s^2 - \sigma_d^2 = \frac{1}{4}\sigma_{01}^2 + \frac{1}{16}\sigma_{02}^2 + \frac{1}{8}\sigma_{11}^2 + \cdots$, is the difference between the covariance among full sibs and the covariances among paternal sibs and among maternal sibs, ($\sigma_{FS}^2 = \sigma_s^2 + \sigma_d^2 + \sigma_{(sd)}^2$ is the covariance among full sibs);

$\sigma_w^2 = \sigma_g^2 + \sigma_e^2 - \sigma_s^2 - \sigma_d^2 - \sigma_{(sd)}^2$, the total phenotypic variance minus σ_{FS}^2.

Note that σ_d^2 of the two-way nested classification model includes both σ_d^2 and $\sigma_{(sd)}^2$ of the two-way cross classification model.

If $\sigma_g^2 = \sigma_{10}^2 + \sigma_{01}^2$, then, $\sigma_{(sd)}^2 = \frac{1}{4}\sigma_{01}^2$ and $\sigma_w^2 = \frac{1}{2}\sigma_{10}^2 + \frac{3}{4}\sigma_{01}^2 + \sigma_e^2$.

In terms of different combinations of the P's the expected values are:

$$E(P_{ijk}) = \mu$$

$$E(P_{ijk}^2) = \mu^2 + \sigma_s^2 + \sigma_d^2 + \sigma_{(sd)}^2 + \sigma_w^2$$

$$E(P_{ijk}P_{ijk'}) = \mu^2 + \sigma_s^2 + \sigma_d^2 + \sigma_{(sd)}^2 \quad \text{(records of full sibs)}$$

$$E(P_{ijk}P_{ij'k'}) = \mu^2 + \sigma_s^2 \quad \text{(records of paternal half sibs)}$$

$$E(P_{ijk}P_{i'jk'}) = \mu^2 + \sigma_d^2 \quad \text{(records of maternal half sibs)}$$

$$E(P_{ijk}P_{i'j'k'}) = \mu^2 \quad \text{(records of unrelated animals)}$$

These definitions of the model lead to the following expectations of the five quadratics used to estimate σ_s^2, σ_d^2, $\sigma_{(sd)}^2$ and σ_w^2 :

$$E\left(\sum_i \sum_j \sum_k P_{ijk}^2\right) = n..\mu^2 + n..\sigma_s^2 + n..\sigma_d^2 + n..\sigma_{(sd)}^2 + n..\sigma_w^2$$

$$E\left(\sum \sum \frac{P_{ij.}^2}{n_{ij}}\right) = n..\mu^2 + n..\sigma_s^2 + n..\sigma_d^2 + n..\sigma_{(sd)}^2 + C\,\sigma_w^2$$

$$E\left(\sum \frac{P_{i..}^2}{n_{i.}}\right) = n..\mu^2 + n..\sigma_s^2 + k_{12}\sigma_d^2 + k_{12}\sigma_{(sd)}^2 + S\,\sigma_w^2$$

$$E\left(\sum \frac{P_{.j.}^2}{n_{.j}}\right) = n..\mu^2 + k_{21}\sigma_s^2 + n..\sigma_d^2 + k_{21}\sigma_{(sd)}^2 + D\,\sigma_w^2$$

$$E\left(\frac{P_{...}^2}{n..}\right) = n..\mu^2 + k_1\,\sigma_s^2 + k_2\,\sigma_d^2 + k_3\,\sigma_{(sd)}^2 + \sigma_w^2 \quad \text{with}$$

$$k_{12} = \sum_i \frac{\sum_j n_{ij}^2}{n_{i.}} \quad , \quad k_{21} = \sum_j \frac{\sum_i n_{ij}^2}{n_{.j}},$$

$$k_1 = \frac{1}{n..} \sum_i n_{i.}^2, \quad k_2 = \frac{1}{n..} \sum_j n_{.j}^2 \quad , \text{and} \quad k_3 = \frac{1}{n..} \sum_i \sum_j n_{ij}^2$$

When the quadratics are equated to their expected values the estimates are:

$$\hat{\sigma}_w^2 = \left(\sum_i \sum_j \sum_k P_{ijk}^2 - \sum_i \sum_j \frac{P_{ij.}^2}{n_{ij}} \right) \; / \; (n.. - C)$$

Let:

$$R_A = \sum_i \frac{P_{i..}^2}{n_{i.}} - \frac{P_{...}^2}{n..} - (S - 1)\, \hat{\sigma}_w^2$$

$$R_B = \sum_j \frac{P_{.j.}^2}{n_{.j}} - \frac{P_{...}^2}{n..} - (D - 1)\, \hat{\sigma}_w^2$$

$$R_{AB} = \sum_i \sum_j \frac{P_{ij.}^2}{n_{ij}} - \sum_i \frac{P_{i..}^2}{n_{i.}} - \sum_j \frac{P_{.j.}^2}{n_{.j}} + \frac{P_{...}^2}{n..} - (C - S - D + 1)\, \hat{\sigma}_w^2$$

Then:

$$\hat{\sigma}_{(sd)}^2 = - \left[\frac{R_A \left(\frac{k_1 - k_{21}}{n.. - k_{21}} \right) + R_B \left(\frac{k_2 - k_{12}}{n.. - k_{12}} \right) - R_{AB} \left(1 - \frac{k_1 - k_{21}}{n.. - k_{21}} - \frac{k_2 - k_{12}}{n.. - k_{12}} \right)}{n.. - k_1 - k_2 + k_3} \right]$$

Also:

$$\hat{\sigma}_s^2 = \left(R_A + R_{AB}\right) \; / \; \left(n.. - k_{21}\right) - \hat{\sigma}_{(sd)}^2$$

$$\hat{\sigma}_d^2 = \left(R_B + R_{AB}\right) \; / \; \left(n.. - k_{12}\right) - \hat{\sigma}_{(sd)}^2$$

$$\hat{\sigma}_P^2 = \hat{\sigma}_s^2 + \hat{\sigma}_d^2 + \hat{\sigma}_{(sd)}^2 + \hat{\sigma}_w^2$$

If $\sigma_g^2 = \sigma_{10}^2 + \sigma_{01}^2$ then there are three estimates of σ_{10}^2 :

$$\hat{\sigma}_{10}^2 = 4\,\hat{\sigma}_s^2 \;\; ; \;\; \hat{\sigma}_{10}^2 = 4\,\hat{\sigma}_d^2 \;\; ; \;\; \hat{\sigma}_{10}^2 = 2\left(\hat{\sigma}_s^2 + \hat{\sigma}_d^2\right).$$

The three corresponding estimates of heritability in the narrow sense are:

$$h_s^2 = \frac{4\,\hat{\sigma}_s^2}{\hat{\sigma}_P^2} \;\; ; \;\; h_d^2 = \frac{4\,\hat{\sigma}_d^2}{\hat{\sigma}_P^2} \;\; ; \;\; h_{(s+d)}^2 = \frac{2\left(\hat{\sigma}_s^2 + \hat{\sigma}_d^2\right)}{\hat{\sigma}_P^2}.$$

The estimate of σ_{01}^2 is: $\hat{\sigma}_{01}^2 = 4\,\hat{\sigma}_{(sd)}^2$

Variations Of Cross-Classified Models

Often the two-way cross classified model is used when herds are considered as random effects and sires have paternal half sib progeny in more than one herd. If the herd effect is substituted for the dam effect in the model the corresponding variance component is due to herd differences and the interaction component is due to sire by herd interaction rather than dominance effects of the sire by dam model.

Another variation of this model is to eliminate the interaction term and estimate σ_s^2, σ_d^2 and σ_w^2 only. The interaction variance, $\sigma_{(sd)}^2$, may be poorly estimated because of few filled subclasses and not many observations in each filled subclass. A method of

estimation for this model is to equate the total, sire, dam, and correction term sums of squares to their expectations which now do not contain $\sigma^2_{(sd)}$. For some traits, the dam component may include variance due to maternal effects, i.e., $\sigma^2_d = \frac{1}{4}\sigma^2_{10} + \frac{1}{16}\sigma^2_{20} + \cdots + \sigma^2_m$ in which case $\hat{\sigma}^2_d$ cannot be used to estimate σ^2_{10} or σ^2_{20}.

Example of Estimating Heritability for the Two-way Cross Classification Model

The following set of 10 progeny records from 3 sires and 4 dams was drawn from a population with $\mu = 50$, $\sigma^2_s = 49$, $\sigma^2_d = 49$, $\sigma^2_{(sd)} = 25$ and $\sigma^2_w = 377$.

Sire (i)	Dam (j)	Progeny (k)	P_{ijk}	Sire (i)	Dam (j)	Progeny (k)	P_{ijk}
1	1	1	135	2	4	1	147
1	1	1	99	2	4	2	140
			$P_{11.} = 234$				$P_{24.} = 287$
1	2	1	153	3	2	1	193
			$P_{12.} = 153$				$P_{32.} = 193$
2	1	1	149	3	3	1	146
			$P_{21.} = 149$	3	3	2	156
							$P_{33.} = 302$
2	3	1	157				
			$P_{23.} = 157$				

One way to set-up the subclass and class totals needed to compute the quadratics is to create a table of subclass totals as shown.

Table of subclass totals, $P_{ij.}$, and numbers, n_{ij}

		Sire (i)				
		1	2	3	$P_{.j.}$	$n_{.j}$
Dam (j)	1	234 (2)	149 (1)	---	383	3
	2	153 (1)	---	193 (1)	346	2
	3	---	157 (1)	302 (2)	459	3
	4	---	287 (2)	---	287	2
	$P_{i..}$	387	593	495	$P_{...} = 1475$	
	$n_{i.}$	3	4	3	$n_{..} = 10$	

$$\sum_i \sum_j \sum_k P_{ijk}^2 = 135^2 + 99^2 + \cdots + 146^2 + 156^2 = 222{,}395$$

$$\sum_i \sum_j \frac{P_{ij.}^2}{n_{ij}} = \frac{234^2}{2} + \frac{153^2}{1} + \cdots + \frac{302^2}{2} = 221{,}672.5$$

$$\sum_i \frac{P_{i..}^2}{n_{i.}} = \frac{387^2}{3} + \frac{593^2}{4} + \frac{495^2}{3} = 219{,}510.25$$

$$\sum_j \frac{P_{.j.}^2}{n_{.j}} = \frac{383^2}{3} + \frac{346^2}{2} + \frac{459^2}{3} + \frac{287^2}{2} = 220{,}165.83$$

$$\frac{P_{...}^2}{n_{..}} = \frac{1475^2}{10} = 217{,}562.5$$

$S = 3$, $D = 4$, and $C = 7$

$$k_1 = \frac{1}{10}\left[3^2 + 4^2 + 3^2\right] = 3.4, \quad k_2 = \frac{1}{10}\left[3^2 + 2^2 + 3^2 + 2^2\right] = 2.6$$

$$k_3 = \frac{1}{10}\left[2^2 + 1^2 + 1^2 + 1^2 + 2^2 + 1^2 + 2^2\right] = 1.6$$

$$k_{12} = \frac{2^2 + 1^2}{3} + \frac{1^2 + 1^2 + 2^2}{4} + \frac{1^2 + 2^2}{3} = 4.83$$

$$k_{21} = \frac{2^2 + 1^2}{3} + \frac{1^2 + 1^2}{2} + \frac{1^2 + 2^2}{3} + \frac{2^2}{2} = 6.33$$

$\hat{\sigma}^2_w$ = (222,395 - 221,672.5)/(10-7) = 240.8

R_A = 219,510.25 - 217,562.5 - (3-1)(240.8) = 1466.15

R_B = 220,165.83 - 217,562.5 - (4-1)(240.8) = 1880.93

R_{AB} = 221,672.5 - 219,510.25 - 220,165.83 + 217,562.5 - (7-3-4+1)(240.8) = -681.88

$(k_1 - k_{21})/(n_{..} - k_{21})$ = (3.40 - 6.33)/(10 - 6.33) = -.7984

$(k_2 - k_{12})/(n_{..} - k_{12})$ = (2.60 - 4.83)/(10 - 4.83) = -.4313

$n_{..} - k_1 - k_2 + k_3 = 5.6$

$\hat{\sigma}^2_{(sd)}$ = -[1466.15(-.7984) + 1880.93(-.4313) - (-681.88)(1 + .7984 + .4313)]/5.6 = 82.4

$\hat{\sigma}^2_s$ = [1466.15 + (-681.88)]/(10 - 6.33) - 82.4 = 131.3

$\hat{\sigma}^2_d$ = [1880.93 + (-681.88)]/(10 - 4.83) - 82.4 = 149.5

$\hat{\sigma}^2_P$ = 131.3 + 149.5 + 82.4 + 240.8 = 604.0

$$h^2_s = \frac{4(131.3)}{604.0} = .87; \quad h^2_d = \frac{4(149.5)}{604.0} = .99; \quad h^2_{(s+d)} = \frac{.87 + .99}{2} = .93$$

$\hat{\sigma}^2_{01}$ = 4(824) = 329.6 $\quad \hat{\sigma}^2_{10(s+d)}$ = 2(131.3 + 149.5) = 561.6

Because of the small sample size these are not reliable estimates, e.g. $\hat{\sigma}^2_{10} + \hat{\sigma}^2_{01} = 891.2$ exceeds $\hat{\sigma}^2_P = 604$. Even with many records the estimate of any component except σ^2_w may be negative.

CHAPTER 39

GENETIC, ENVIRONMENTAL, AND PHENOTYPIC CORRELATIONS

If X and Y denote the two traits then the genetic, environmental and phenotypic correlations are defined as:

genetic,
$$r_g = \frac{\sigma_{G_x G_y}}{\sqrt{\sigma^2_{G_x}\ \sigma^2_{G_y}}} \ ;$$

environmental,
$$r_e = \frac{\sigma_{E_x E_y}}{\sqrt{\sigma^2_{E_x}\ \sigma^2_{E_y}}} \ ; \text{ and}$$

phenotypic,
$$r_p = \frac{\sigma_{P_x P_y}}{\sqrt{\sigma^2_{P_x}\ \sigma^2_{P_y}}} = \frac{\sigma_{G_x G_y} + \sigma_{E_x E_y}}{\sqrt{\left(\sigma^2_{G_x} + \sigma^2_{E_x}\right)\left(\sigma^2_{G_y} + \sigma^2_{E_y}\right)}}$$

where σ^2_G, σ^2_E, and σ^2_P are genetic, environmental, and phenotypic variances and $\sigma_{G_x G_y}$, $\sigma_{E_x E_y}$ and $\sigma_{P_x P_y}$ are the corresponding genetic, environmental and phenotypic covariances between traits X and Y. Usually only additive genetic effects are assumed to contribute to the genetic variances and covariances.

ESTIMATION FROM PARENT AND OFFSPRING COVARIANCE

This method is comparable to regression and correlation methods of estimating heritability. The model is a two-trait version of the one used in Chapter 38 (see also Figure 39.1 for diagram of the traits and their variances and covariances):

$$\text{Parent, p, Trait x: } P_{ipx} = \mu_x + G_{ipx} + E_{ipx}$$

$$\text{Parent, p, Trait y: } P_{ipy} = \mu_y + G_{ipy} + E_{ipy}$$

$$\text{Offspring, o, Trait x: } P_{iox} = \mu_x + G_{iox} + E_{iox}$$

$$\text{Offspring, o, Trait y: } P_{ioy} = \mu_y + G_{ioy} + E_{ioy}$$

with i=1, ..., n where n is the number of sets of the four observations. A basic assumption is that all G's are additive genetic values and E's are random environmental effects.

The G's and E's are $\text{IID}\left(0,\ \sigma^2_G \text{ and } \sigma^2_E\right)$ as before but now records on traits X and Y on the same animal have covariance, $\sigma_{P_x P_y} = \sigma_{G_x G_y} + \sigma_{E_x E_y}$; records of the same traits on parent and offspring have covariance (1/2) σ^2_G, and records with one trait on the parent and the other trait on the offspring have covariance (1/2) $\sigma_{G_x G_y}$. In terms of expectations of the P's:

$$E(P_{ipx}) = E(P_{iox}) = \mu_x \ ; \ E(P_{ipy}) = E(P_{ioy}) = \mu_y$$

$$E\left(P^2_{ipx}\right) = E\left(P^2_{iox}\right) = \mu^2_x + \sigma^2_{G_x} + \sigma^2_{E_x}$$

$$E\left(P^2_{ipy}\right) = E\left(P^2_{ioy}\right) = \mu^2_y + \sigma^2_{G_y} + \sigma^2_{E_y}$$

$$E(P_{ipx}P_{i'px}) = E(P_{iox}P_{i'ox}) = E(P_{ipx}P_{i'ox}) = \mu^2_x$$

$$E(P_{ipy}P_{i'py}) = E(P_{ioy}P_{i'oy}) = E(P_{ipy}P_{i'\sigma y}) = \mu^2_y$$

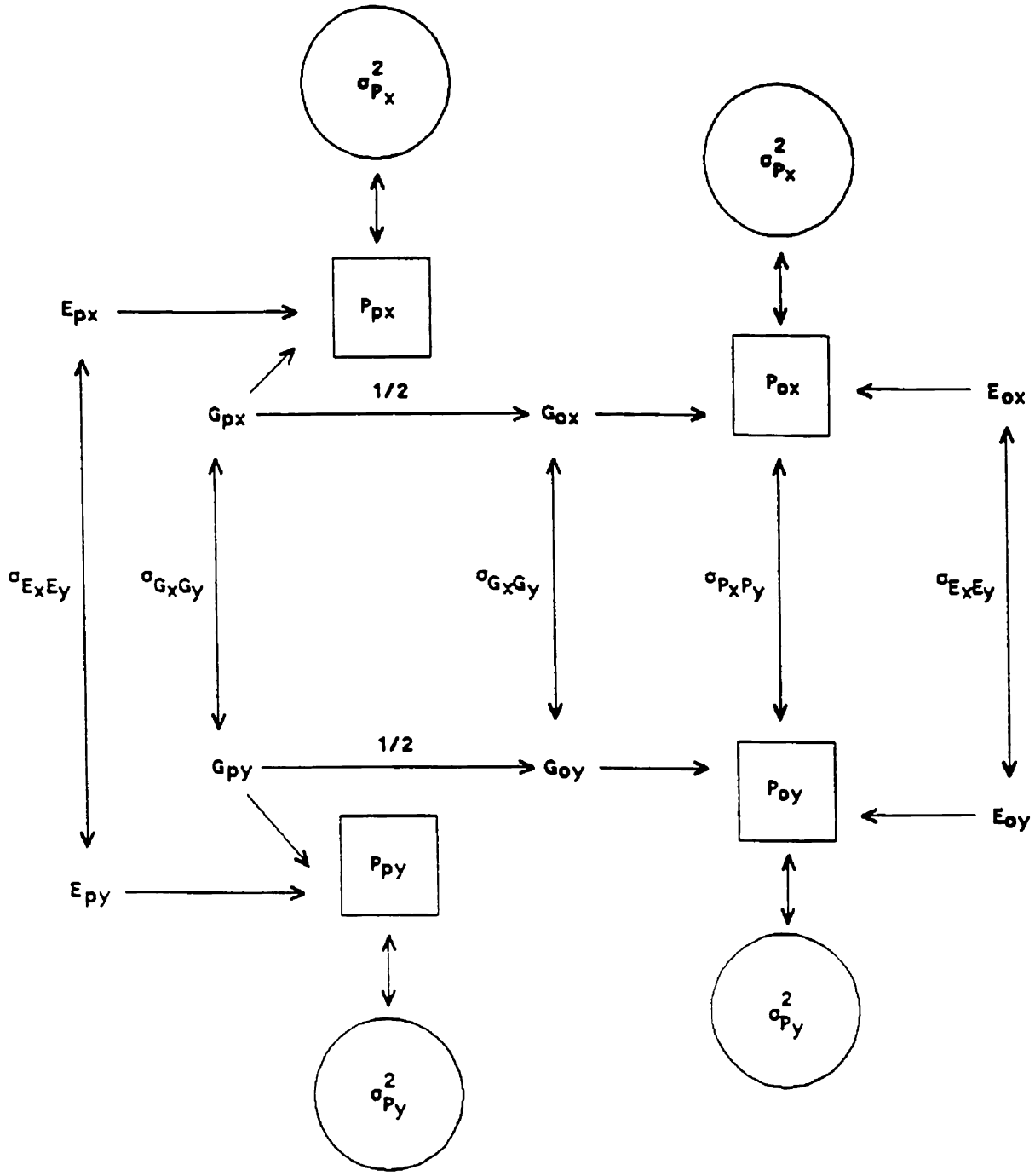


Figure 38.1. Diagram of the variable effects in the model for parent-offspring records on traits x and y. Only the phenotypic records (boxes) are observed. Circled phenotypic variances can be computed from the phenotypic records. Phenotypic covariances can be estimated from (P_{px} , P_{py}) and from (P_{ox} , P_{oy}).

$$E(P_{ipx}P_{iox}) = \mu_x^2 + \frac{1}{2}\sigma^2_{G_x}; E(P_{ipy}P_{ioy}) = \mu_y^2 + \frac{1}{2}\sigma^2_{G_y}$$

$$E(P_{ipx}P_{i'ox}) = \mu_x^2 \; ; \; E(P_{ipy}P_{i'oy}) = \mu_y^2$$

$$E(P_{ipx}P_{ioy}) = E(P_{iox}P_{ipy}) = \mu_x\mu_y + \frac{1}{2}\sigma_{G_xG_y}$$

$$E(P_{ipx}P_{i'oy}) = E(P_{iox}P_{i'py}) = \mu_x\mu_y$$

$$E(P_{ipx}P_{ipy}) = E(P_{iox}P_{ioy}) = \mu_x\mu_y + \sigma_{G_xG_y} + \sigma_{E_xE_y}$$

$$E(P_{ipx}P_{i'py}) = E(P_{iox}P_{i'oy}) = \mu_x\mu_y$$

The method of estimation consists of computing the variances and covariances among the four phenotypic measurements for the parent-offspring sets. The expectations of the total sums of squares and products and of the correction terms are shown on the next page.

E[Total Sums of Squares and Products]

(all times n)

$\sum_i$	P_{ipx}	P_{iox}	P_{ipy}	P_{ioy}
P_{ipx}	$\mu_x^2 + \sigma^2_{G_x} + \sigma^2_{E_x}$	$\mu_x^2 + .5\,\sigma^2_{G_x}$	$\mu_x\mu_y + \sigma_{G_x G_y} + \sigma_{E_x E_y}$	$\mu_x\mu_y + .5\,\sigma_{G_x G_y}$
P_{iox}		$\mu_x^2 + \sigma^2_{G_x} + \sigma^2_{E_x}$	$\mu_x\mu_y + .5\,\sigma_{G_x G_y}$	$\mu_x\mu_y + \sigma_{G_x G_y} + \sigma_{E_x E_y}$
P_{ipy}			$\mu_y^2 + \sigma^2_{G_y} + \sigma^2_{E_y}$	$\mu_y^2 + .5\,\sigma^2_{G_y}$
P_{ioy}	symmetric			$\mu_y^2 + \sigma^2_{G_y} + \sigma^2_{E_y}$

E[Correction Factors]

	$P_{.px}$	$P_{.ox}$	$P_{.py}$	$P_{.oy}$
$P_{.px}$	$n\mu_x^2 + \sigma^2_{G_x} + \sigma^2_{E_x}$	$n\mu_x^2 + .5\,\sigma^2_{G_x}$	$n\mu_x\mu_y + \sigma_{G_x G_y} + \sigma_{E_x E_y}$	$n\mu_x\mu_y + .5\,\sigma_{G_x G_y}$
$P_{.ox}$		$n\mu_x^2 + \sigma^2_{G_x} + \sigma^2_{E_x}$	$n\mu_x\mu_y + .5\,\sigma_{G_x G_y}$	$n\mu_x\mu_y + \sigma_{G_x G_y}$
$P_{.py}$			$n\mu_y^2 + \sigma^2_{G_y} + \sigma^2_{E_y}$	$n\mu_x\mu_y + .5\,\sigma^2_{G_y}$
$P_{.oy}$	symmetric			$n\mu_y^2 + \sigma^2_{G_y} + \sigma^2_{E_y}$

The four variances are computed as usual as:

$$\hat{\sigma}^2_u = \left(\sum_i P^2_{iu} - \frac{P^2_{.u}}{n} \right) / (n-1) \text{ where [u = px, ox, py, and oy].}$$

The six covariances are computed as:

$$\hat{\sigma}_{uv} = \left(\sum_i P_{iu} P_{iv} - \frac{P_{.u} P_{.v}}{n} \right) / (n-1)$$

where [(u,v) = (px,ox), (px,py), (px,oy),(ox,py), (ox,oy), and (py,oy)].

The expected values of these four variances and covariances as summarized below are then used to estimate the genetic, environmental and phenotypic covariances and correlations.

Table of expected values of estimated variances and covariances

(Diagonals are variances, off-diagonals are covariances.)

	P_{px}	P_{ox}	P_{py}	P_{oy}
P_{px}	$\sigma^2_{G_x} + \sigma^2_{E_x}$	$.5\ \sigma^2_{G_x}$	$\sigma_{G_xG_y} + \sigma_{E_xE_y}$	$.5\ \sigma_{G_xG_y}$
P_{ox}		$\sigma^2_{G_x} + \sigma^2_{E_x}$	$.5\ \sigma_{G_xG_y}$	$\sigma_{G_xG_y} + \sigma_{E_xE_y}$
P_{py}			$\sigma^2_{G_y} + \sigma^2_{E_y}$	$.5\ \sigma^2_{G_y}$
P_{oy}	symmetric			$\sigma^2_{G_y} + \sigma^2_{E_y}$

There are several ways of combining these quantities. The usual ways are given here.

Four Possible Estimates of r_g

Let $D = \sqrt{\hat{\sigma}_{px,ox}\ \hat{\sigma}_{py,oy}}$.

1) $r_g = \hat{\sigma}_{px,oy} / D$

2) $r_g = \hat{\sigma}_{ox,py} / D$

3) $r_g = .5\left(\hat{\sigma}_{px,oy} + \hat{\sigma}_{ox,py}\right) / D$ (the arithmetic average of 1 and 2) and

4) $r_g = \sqrt{\hat{\sigma}_{px,oy}\ \hat{\sigma}_{ox,py}} / D$ (the geometric mean of 1 and 2).

Note:

a) The expected value by parts for each of the four estimates is:

$$r_g = \frac{.5\ \sigma_{G_x G_y}}{\sqrt{.5\ \sigma^2_{G_x}\ .5\ \sigma^2_{G_y}}} = \frac{\sigma_{G_x G_y}}{\sqrt{\sigma^2_{G_x}\ \sigma^2_{G_y}}} .$$

b) Estimates must be discarded if either or both covariances in D is negative (i.e., these covariances correspond to negative heritabilities).

c) Estimates by method 4) must also be discarded if the signs of covariances in the numerator are different.

d) If both covariances in the numerator of 4) are negative, assign a negative sign to r_g.

The sampling variances of 3) and 4) are equal and are approximately one-fourth of those for 1) and 2). Method 3 probably is best since fewer estimates will be discarded than with method 4.

Two Estimates of r_p

1) From parents:

$$\hat{r}_p = \frac{\hat{\sigma}_{px,py}}{\sqrt{\hat{\sigma}^2_{px}\,\hat{\sigma}^2_{py}}}$$

2) From offspring:

$$\hat{r}_p = \frac{\hat{\sigma}_{ox,oy}}{\sqrt{\hat{\sigma}^2_{ox}\,\hat{\sigma}^2_{oy}}}$$

The arithmetic and geometric means of 1) and 2) can also be used. The expected values by parts are: $\dfrac{\sigma_{P_x P_y}}{\sqrt{\sigma^2_{P_x}\,\sigma^2_{P_y}}}$ for any of these estimates .

Estimate of r_e

$$r_e = \frac{\hat{\sigma}_{px,py} + \hat{\sigma}_{ox,oy} - 2\left(\hat{\sigma}_{px,oy} + \hat{\sigma}_{ox,py}\right)}{\sqrt{\left(\hat{\sigma}^2_{px} + \hat{\sigma}^2_{ox} - 4\,\hat{\sigma}_{px,ox}\right)\left(\hat{\sigma}^2_{py} + \hat{\sigma}^2_{oy} - 4\,\hat{\sigma}_{py,oy}\right)}}$$

Note as in a previous section on estimating heritability by regression and correlation:

$$h^2_x = \frac{2\,\hat{\sigma}_{px,ox}}{\hat{\sigma}^2_{px}} \quad \text{or} \quad h^2_x = \frac{2\,\hat{\sigma}_{px,ox}}{\sqrt{\hat{\sigma}^2_{px}\,\hat{\sigma}^2_{ox}}} \quad \text{and}$$

$$h^2_y = \frac{2\,\hat{\sigma}_{py,oy}}{\hat{\sigma}^2_{py}} \quad \text{or} \quad h^2_y = \frac{2\,\hat{\sigma}_{py,oy}}{\sqrt{\hat{\sigma}^2_{py}\,\hat{\sigma}^2_{oy}}}$$

ESTIMATION FROM COMPONENTS OF VARIANCES AND COVARIANCES

One-Way Classification (extension to more complex classification models is similar)

The model for the two trait, one-way classification model is analogous to the single trait models used for estimating heritability with sires providing the group classification:

Trait x: $P_{xij} = \mu_x + b_{xi} + w_{xij}$

Trait y: $P_{yij} = \mu_y + b_{yi} + w_{yij}$ both measured on animal ij, where

μ is a constant for the appropriate trait,

b_i is an effect common to $i^{\underline{th}}$ genetic group for the appropriate trait,

w_{ij} is an effect associated with $j^{\underline{th}}$ member of $i^{\underline{th}}$ group,

a_{ii} is the additive relationship among animals in group i with

$i = 1, ..., B;\ j = 1, ..., n_i$.

If all genetic effects are additive genetic effects then:

$$\sigma^2_{b_x} = a_{ii}\,\sigma^2_{G_x}\ ;\ \sigma^2_{b_y} = a_{ii}\,\sigma^2_{G_y}\ ;\ \text{and}\ \ \sigma_{b_x b_y} = a_{ii}\,\sigma_{G_x G_y}$$

$$\sigma^2_{w_x} = (1 - a_{ii})\,\sigma^2_{G_x} + \sigma^2_{E_x}\ ;\ \sigma^2_{w_y} = (1 - a_{ii})\,\sigma^2_{G_y} + \sigma^2_{E_y}\ ;\ \text{and}$$

$$\sigma_{w_x w_y} = (1 - a_{ii})\,\sigma_{G_x G_y} + \sigma_{E_x E_y}$$

The b's are $\text{IID}\left(0,\ \sigma^2_{b_x}\ \text{or}\ \sigma^2_{b_y}\right)$, the w's are $\text{IID}\left(0,\ \sigma^2_{w_x}\ \text{or}\ \sigma^2_{w_y}\right)$, the b_x's are uncorrelated with w_x's, the b_y's and w_y's are uncorrelated, b_{xi} and b_{yi} have covariance $\sigma_{b_x b_y} = a_{ii}\,\sigma_{G_x G_y}$; and w_{xij} and w_{yij} (on the same animal) have covariance, $\sigma_{P_x P_y} - \sigma_{b_x b_y}$.

In terms of the b's these definitions are:

$$E(P_{xij}) = \mu_x \;; \qquad E(P_{yij}) = \mu_y$$

$$E(P_{xij}^2) = \mu_x^2 + \sigma_{b_x}^2 + \sigma_{w_x}^2 \;; \qquad E(P_{yij}^2) = \mu_y^2 + \sigma_{b_y}^2 + \sigma_{w_y}^2$$

$$E(P_{xij}P_{yij}) = \mu_x\mu_y + \sigma_{b_x b_y} + \sigma_{w_x w_y}$$

$$E(P_{xij}P_{xij'}) = \mu_x^2 + \sigma_{b_x}^2 \;; \quad E(P_{yij}P_{yij'}) = \mu_y^2 + \sigma_{b_y}^2; \quad E(P_{xij}P_{yij'}) = \mu_x\mu_y + \sigma_{b_x b_y}$$

$$E(P_{xij}P_{xi'j'}) = \mu_x^2 \;; \qquad E(P_{yij}P_{yi'j'}) = \mu_y^2; \qquad E(P_{xij}P_{yi'j'}) = \mu_x\mu_y$$

These definitions lead to the following expectations of the six sums of squares and three sums of products used for estimating the variances and covariances.

$$E\left(\sum_i \sum_j P_{xij}^2\right) = n.\mu_x^2 + n.\sigma_{b_x}^2 + n.\sigma_{w_x}^2$$

$$E\left(\sum_i \frac{P_{xi.}^2}{n_i}\right) = n.\mu_x^2 + n.\sigma_{b_x}^2 + B\,\sigma_{w_x}^2$$

$$E\left(\frac{P_{x..}^2}{n.}\right) = n.\mu_x^2 + \frac{1}{n.}\sum_i n_i^2\sigma_{b_x}^2 + \sigma_{w_x}^2$$

$$E\left(\sum_i \sum_j P_{yij}^2\right) = n.\mu_y^2 + n.\sigma_{b_y}^2 + n.\sigma_{w_y}^2$$

$$E\left(\sum_i \frac{P_{yi.}^2}{n_i}\right) = n.\mu_y^2 + n.\sigma_{b_y}^2 + B\,\sigma_{w_y}^2$$

$$E\left(\frac{P_{y..}^2}{n.}\right) = n.\mu_y^2 + \frac{1}{n.}\sum_i n_i^2\sigma_{b_y}^2 + \sigma_{w_y}^2$$

$$E\left(\sum_i \sum_j P_{xij} P_{yij}\right) = n.\mu_x\mu_y + n.\sigma_{b_x b_y} + n.\sigma_{w_x w_y}$$

$$E\left(\sum \frac{P_{xi.}P_{yi.}}{n_i}\right) = n.\mu_x\mu_y + n.\sigma_{b_x b_y} + B\,\sigma_{w_x w_y}$$

$$E\left(\frac{P_{x..}P_{y..}}{n.}\right) = n.\mu_x\mu_y + \frac{1}{n.}\sum_i n_i^2 \sigma_{b_x b_y} + \sigma_{w_x w_y}$$

The expectations of the quadratics are equated to their computed values to estimate the variances and covariances. Estimates of the variance components are as before:

$$\hat{\sigma}^2_{w_u} = \left(\sum_i \sum_j P^2_{uij} - \sum_i \frac{P^2_{ui.}}{n_i}\right) / (n. - B)$$

$$\hat{\sigma}^2_{b_u} = \left(\sum_i \frac{P^2_{ui.}}{n_i} - \frac{P^2_{u..}}{n.} - (B-1)\,\hat{\sigma}^2_w\right) / \left(n. - \frac{1}{n.}\sum_i n_i^2\right) \text{ and}$$

$$\hat{\sigma}^2_{p_u} = \hat{\sigma}^2_{b_u} + \hat{\sigma}^2_{w_u} \text{ where } u = x \text{ or } y$$

Similarly the estimates of the covariance components are:

$$\hat{\sigma}_{w_x w_y} = \left(\sum_i \sum_j P_{xij} P_{yij} - \sum_i \frac{P_{xi.}\,P_{yi.}}{n_i}\right) / (n. - B)$$

$$\hat{\sigma}_{b_x b_y} = \left(\sum_i \frac{P_{xi.}\,P_{yi.}}{n_i} - \frac{P_{x..}\,P_{y..}}{n.} - (B-1)\,\hat{\sigma}_{w_x w_y}\right) / \left(n. - \frac{1}{n.}\sum_i n_i^2\right)$$

$$\hat{\sigma}_{p_x p_y} = \hat{\sigma}_{b_x b_y} + \hat{\sigma}_{w_x w_y}$$

Then,

$$r_g = \hat{\sigma}_{b_x b_y} / \sqrt{\hat{\sigma}^2_{b_x}\,\hat{\sigma}^2_{b_y}} \quad ; \qquad r_p = \hat{\sigma}_{p_x p_y} / \sqrt{\hat{\sigma}^2_{p_x}\,\hat{\sigma}^2_{p_y}} \quad ; \quad \text{and}$$

$$r_e = \frac{\hat{\sigma}_{w_x w_y} - \left(\frac{1-a_{ii}}{a_{ii}}\right)\hat{\sigma}_{b_x b_y}}{\sqrt{\left[\hat{\sigma}^2_{w_x} - \left(\frac{1-a_{ii}}{a_{ii}}\right)\hat{\sigma}^2_{b_x}\right]\left[\hat{\sigma}^2_{w_y} - \left(\frac{1-a_{ii}}{a_{ii}}\right)\hat{\sigma}^2_{b_y}\right]}} .$$

Estimates of variance components for the one-way classification model provide, as before, estimates of heritability:

$$h^2_x = \frac{1}{a_{ii}}\,\hat{\sigma}^2_{b_x}/\hat{\sigma}^2_{p_x} \quad \text{and} \quad h^2_y = \frac{1}{a_{ii}}\,\hat{\sigma}^2_{b_y}/\hat{\sigma}^2_{p_y} .$$

The expectations by parts of these estimates are r_g, r_p, r_e, h^2_x and h^2_y. The estimates of heritability can be outside the parameter limits of 0 and 1. The estimate of genetic correlation may be, and with small data sets often is, outside the parameter limits of -1 and +1.

Example of Estimating Genetic, Phenotypic and Environmental Correlations

In this example of genetic groups of paternal half-sibs, each has one record.

The following set of 75 records from 25 groups was drawn from a population with $\mu_x = 1500$, $\mu_y = 200$, $\sigma^2_{b_x} = 2500$, $\sigma^2_{b_y} = 225$, $\sigma_{b_x b_y} = 375$, $\sigma^2_{w_x} = 47{,}500$, $\sigma^2_{w_y} = 1{,}575$, and $\sigma_{w_x w_y} = 4{,}125$ which correspond to $h^2_x = .20$, $h^2_y = .50$, $r_g = .50$, $r_e = .50$ and $r_p = .47$.

i	j	P_{xij}	P_{yij}
1	1	1878	264
1	2	1542	289
1	.	3420	553
2	1	1493	151
2	2	1598	170
2	3	1727	176
2	.	4818	497
3	1	1371	190
3	.	1371	190
4	1	1453	173
4	2	1588	220
4	3	1207	146
4	4	1673	278
4	5	1373	217
4	.	7294	1034
5	1	1612	173
5	2	1402	172
5	3	1389	141
5	4	958	127
5	.	5361	613
6	1	1444	230
6	2	1719	184
6	.	3163	414
7	1	1552	212
7	2	1327	206
7	3	1364	179
7	.	4243	597

i	j	P_{xij}	P_{yij}
8	1	1499	211
8	.	1499	211
9	1	1328	208
9	2	1887	240
9	3	1498	154
9	4	1920	181
9	5	1785	169
9	.	8418	952
10	1	1854	238
10	2	1043	165
10	3	1509	91
10	4	1298	182
10	.	5704	676
11	1	1537	218
11	2	1894	223
11	.	3431	441
12	1	1241	180
12	2	1204	176
12	3	1534	156
12	.	3979	512
13	1	1418	142
13	.	1418	142
14	1	1667	264
14	2	1613	230
14	3	1285	224
14	4	1404	230
14	5	1841	229
14	.	7810	1177

i	j	P_{xij}	P_{yij}
15	1	1800	232
15	2	1342	216
15	3	1394	243
15	4	1339	175
15	.	5875	866
16	1	1800	198
16	2	1387	182
16	.	3187	380
17	1	1016	165
17	2	1324	229
17	3	1494	188
17	.	3834	582
18	1	1692	213
18	.	1692	213
19	1	1177	134
19	2	1319	177
19	3	1928	260
19	4	1308	205
19	5	1262	172
19	.	6994	948
20	1	1555	182
20	2	1775	189
20	3	1428	220
20	4	1597	172
20	.	6355	763

i	j	P_{xij}	P_{yij}
21	1	1375	202
21	2	1752	162
21	.	3127	364
22	1	1383	225
22	2	1595	189
22	3	1459	180
22	.	4437	594
23	1	1489	220
23	.	1489	220
24	1	1553	200
24	2	1527	188
24	3	1094	185
24	4	1410	178
24	5	1272	166
24	.	6856	917
25	1	1657	160
25	2	1695	218
25	3	1722	236
25	4	1420	199
25	.	6494	813
.	.	**112,269**	**14,669**

$$\Sigma \Sigma P_{xij}^2 = 1878^2 + 1542^2 + 1493^2 + \cdots + 1420^2 = 171{,}709{,}893 \ ;$$

$$\Sigma \Sigma P_{yij}^2 = 264^2 + 289^2 + 151^2 + \cdots + 199^2 = 2{,}966{,}259 \ ;$$

$$\Sigma \Sigma P_{xij} P_{yij} = (1878)(264) + (1542)(289) + \cdots + (1420)(199) = 22{,}201{,}100 \ .$$

$$\sum_i \frac{P_{xi.}^2}{n_i} = \frac{3420^2}{2} + \frac{4818^2}{3} + \cdots + \frac{6494^2}{4} = 169{,}183{,}456.7 \ ;$$

$$\sum_i \frac{P_{yi.}^2}{n_i} = \frac{553^2}{2} + \frac{497^2}{3} + \cdots + \frac{813^2}{4} = 2{,}914{,}421.8 \ ;$$

$$\sum_i \frac{P_{xi.} P_{yi.}}{n_i} = \frac{(3420)(553)}{2} + \frac{(4818)(497)}{3} + \cdots + \frac{(6494)(813)}{4} = 22{,}064{,}326.9 \ ;$$

$$\frac{P_{x..}^2}{n.} = \frac{(112269)^2}{75} = 168{,}057{,}711.5 \ ;$$

$$\frac{P_{y..}^2}{n.} = (14669)^2/75 = 2{,}869{,}060.8 \ ;$$

$$\frac{P_{x..} P_{y..}}{n.} = \frac{(112{,}269)(14669)}{75} = 21{,}958{,}319.5 \ ;$$

$$n. = 75, \ B = 25, \ \Sigma n_i^2 = 2^2 + 3^2 + 1^2 + 5^2 + \ldots + 4^2 = 275 \ .$$

The estimates of variance and covariance components are:

$$\hat{\sigma}_{w_x}^2 = (171{,}709{,}893 - 169{,}183{,}456.7)/(75 - 25) = 50{,}528.7$$

$$\hat{\sigma}_{w_y}^2 = (2{,}966{,}259 - 2{,}914{,}421.8)/(75 - 25) = 1036.7$$

$$\hat{\sigma}_{w_x w_y} = (22{,}201{,}100 - 22{,}064{,}326.9)/(75 - 25) = 2735.5$$

$$\hat{\sigma}^2_{b_x} = [169,183,456.7 - 168,057,711.5 - (25\text{-}1)(50,528.7)]/(75 - 275/75) = -1218.8$$

$$\hat{\sigma}^2_{b_y} = [2,914,421.8 - 2,869,060.8 - (25\text{-}1)(1036.7)]/(75\text{-}275/75) = 287.1$$

$$\hat{\sigma}_{b_x b_y} = [22,064,326.9 - 21,958,319.5 - (25\text{-}1)(2735.5)]/(75\text{-}275/75) = 565.7$$

The estimates of the correlations and heritabilities are:

$$r_g = \frac{565.7}{\sqrt{(-1218.8)(287.1)}}$$ (Imaginary estimate since $\hat{\sigma}^2_{b_x}$ is negative.)

$$r_p = \frac{(565.7 + 2735.5)}{\sqrt{(-1218.8 + 50,528.7)(287.1 + 1036.7)}} = .41$$

$$r_e = \frac{2735.5 - [(1 - .25)/(.25)](565.7)}{\sqrt{[50,528.7 - 3(-1218.8)][1036.7 - 3(287.1)]}} = .34$$

$$h^2_x = \frac{4(-1218.8)}{(-1218.8 + 50,528.7)} = -0.10$$ (An estimate outside the lower limit.)

$$h^2_y = \frac{4(287.1)}{287.1 + 1036.7} = .87 \quad .$$

CHAPTER 40

MONTE CARLO SIMULATION

Simulation of biological models often aids in understanding the simpler models and in building more complex models. Often simulation is a first step in finding out what happens when the usual assumptions are not fulfilled, e.g., when the data are selected. However, simulation has been used in PART III to aid in understanding random chance associated with a set of data and the effects of random sampling on estimates of parameters from relatively small amounts of data.

The procedures described in this chapter will simulate normal and multivariate normal distributions. The method of simulation is to obtain in some way pseudo-random values (also to be called random normal deviates) from a normal distribution with mean, zero, and variance, one. The pseudo-random standard normal values when multiplied by a constant, σ, result in variables with mean, zero, and variance, σ^2. Adding a constant μ results in a variable with mean, μ, and variance, σ^2. For example, suppose v_i is such a random variable from a distribution with mean, zero, and standard deviation, one.

Then $E(v_i\sigma) = \sigma E(v_i) = 0$; $E(\mu + v_i\sigma) = E(\mu) + E(v_i\sigma) = \mu$;

$$E[(v_i\sigma)^2] = E(v_i^2\sigma^2) = \sigma^2 E(v_i^2) = \sigma^2 \quad ; \quad E[(\mu + v_i\sigma - \mu)^2] = \sigma^2.$$

In the description of how to simulate the models described in chapters 37, 38 and 39, the lower case, primed and subscripted letters will represent normal deviates (0,1).

REPEATABILITY MODEL

$$P_{ij} = \mu + A_i + E_{ij}$$

For each animal, i, draw a_i' and e_{ij}', j=1, ..., n_i random normal values and multiply as shown so that $P_{ij} = \mu + a_i'\sigma_A + e_{ij}'\sigma_E$. Thus the A_i will have zero mean and variance, σ_A^2, and the E_{ij} will have zero mean and variance, σ_E^2. The P_{ij} will have mean, μ, and variance, $\sigma_P^2 = \sigma_A^2 + \sigma_E^2$.

Example for animal 1 with n_1 records:

$$P_{11} = \mu + a_1'\sigma_A + e_{11}'\sigma_E$$

$$P_{12} = \mu + a_1'\sigma_A + e_{12}'\sigma_E$$

$$\vdots$$

$$P_{1n_1} = \mu + a_1'\sigma_A + e_{1n_1}'\sigma_E$$

Note that the n_i+1 random normal deviates are independently drawn so that the expectations of products of different combinations of A's and E's are zero.

HERITABILITY MODEL

Regression And Correlation

P_{xi}, P_{yi} are the $i^{\underline{th}}$ pair of records on relatives x and y.

$$P_{xi} = \mu + g_{xi}'\sigma_G + e_{xi}'\sigma_E \quad \text{and} \quad P_{yi} = \mu + a_{xy}g_{xi}'\sigma_G + g_{yi}'\sqrt{1-a_{xy}^2}\,\sigma_G + e_{yi}'\sigma_E \ .$$

This model simulates statistically the genetic model but does not directly mimic Mendelian sampling. The second **g** term in P_{yi} simulates Mendelian sampling due to segregation and recombination and maintains the genetic variance as σ_G^2 as can be seen by taking the expected value, $E[(P_{yi} - \mu)^2]$.

Variance Components

One-way Classification Model

$$P_{ij} = \mu + b_i'\sigma_b + w_{ij}'\sigma_w$$

Note that $\sigma_b^2 = a_{ii}\sigma_G^2$ and $\sigma_w^2 = \sigma_E^2 + \sqrt{1-a_{ii}^2}\ \sigma_G^2$ where a_{ii} is the additive relationship between animals in the i^{th} group. For paternal half-sib groups, $a_{ii} = 1/4$.

Two-way Nested Classification (Dams Within Sires) Model

$$P_{ijk} = \mu + s_i'\sigma_s + d_{ij}'\sigma_d + w_{ijk}'\sigma_w \quad \text{where}$$

$$\sigma_s^2 = (1/4)\,\sigma_{10}^2 \ ; \ \sigma_d^2 = (1/4)\,\sigma_{10}^2 + (1/4)\sigma_{01}^2 \quad \text{(if dominance is included);}$$

$$\sigma_w^2 = \sigma_E^2 - \sigma_s^2 - \sigma_d^2$$

Two-way Nested Model with Repeatability

$$P_{ijk} = \mu + s_i'\sigma_s + c_{ij}'\sigma_c + w_{ijk}'\sigma_w \quad \text{where}$$

$$\sigma_s^2 = (1/4)\,\sigma_G^2 \ ; \ \sigma_c^2 = \sigma_A^2 - \sigma_s^2 \ ; \ \sigma_w^2 = \sigma_E^2$$

Two-way Cross Classification Model

$$P_{ijk} = \mu + s_i'\sigma_s + d_j'\sigma_d + (sd)_{ij}'\sigma_{(sd)} + w_{ijk}'\sigma_w \quad \text{where}$$

$$\sigma_s^2 = (1/4)\,\sigma_{10}^2 \ ; \ \sigma_d^2 = (1/4)\,\sigma_{10}^2 \ ;$$

$$\sigma_{(sd)}^2 = (1/4)\sigma_{01}^2 \quad \text{(if dominance effects are included);}$$

$$\sigma_w^2 = \sigma_G^2 + \sigma_E^2 - \sigma_s^2 - \sigma_d^2 - \sigma_{(sd)}^2$$

GENETIC, ENVIRONMENTAL, AND PHENOTYPIC CORRELATIONS

Parent-Offspring Covariances

$$P_{ipx} = \mu_x + g'_{i1}G_1 + \qquad + e'_{i1}E_1$$

$$P_{ipy} = \mu_y + g'_{i1}G_2 + g'_{i2}G_3 \qquad + e'_{i1}E_2 + e'_{i2}E_3$$

$$P_{iox} = \mu_x + g'_{i1}\frac{G_1}{2} + g'_{i3}\sqrt{\frac{3}{4}}\ G_1 + e'_{i3}E_1$$

$$P_{ioy} = \mu_y + g'_{i1}\frac{G_2}{2} + g'_{i2}\frac{G_3}{2} + g'_{i3}\sqrt{\frac{3}{4}}\ G_2 + g'_{i4}\sqrt{\frac{3}{4}}\ G_3 + e'_{i3}E_2 + e'_{i4}E_3$$

where, $G_1 = \sigma_{G_x}$; $G_2 = \dfrac{\sigma_{G_{xy}}}{\sigma_{G_x}}$; $E_1 = \sigma_{E_x}$; $E_2 = \dfrac{\sigma_{E_{xy}}}{\sigma_{E_x}}$;

$$G_3 = \sqrt{\sigma^2_{G_y} - \left(\frac{\sigma_{G_{xy}}}{\sigma_{G_x}}\right)^2}\ , \qquad E_3 = \sqrt{\sigma^2_{E_y} - \left(\frac{\sigma_{E_{xy}}}{\sigma_{E_x}}\right)^2}$$

These constants allow simulation of a four-variate distribution with the genetic and environmental variances and covariances shown in the description of the model.

Variance And Covariance Components

One-way Classification Model

$$P_{xij} = \mu_x + b'_{xi}\sigma_{b_x} + w'_{xij}\sigma_{w_x}$$

$$P_{yij} = \mu_y + b'_{xi}\frac{\sigma_{b_{xy}}}{\sigma_{b_x}} + b'_{yi}\sqrt{\sigma^2_{b_y} - \left(\frac{\sigma_{b_{xy}}}{\sigma_{b_x}}\right)^2} + w'_{xij}\frac{\sigma_{w_{xy}}}{\sigma_{w_x}} + w'_{yij}\sqrt{\sigma^2_{w_y} - \left(\frac{\sigma_{w_{xy}}}{\sigma_{w_x}}\right)^2}$$

where, $\sigma^2_{b_x} = a_{ii}\sigma^2_{G_x}$; $\sigma^2_{b_y} = a_{ii}\sigma^2_{G_y}$; $\sigma_{b_{xy}} = a_{ii}\sigma_{G_{xy}}$

$$\sigma^2_{w_x} = (1-a_{ii})\sigma^2_{G_x} + \sigma^2_{E_x}\ ;\ \sigma^2_{w_y} = (1-a_{ii})\sigma^2_{G_y} + \sigma^2_{E_y}\ ;\ \sigma_{w_{xy}} = (1-a_{ii})\sigma_{G_{xy}} + \sigma_{E_{xy}}$$

CHAPTER 41

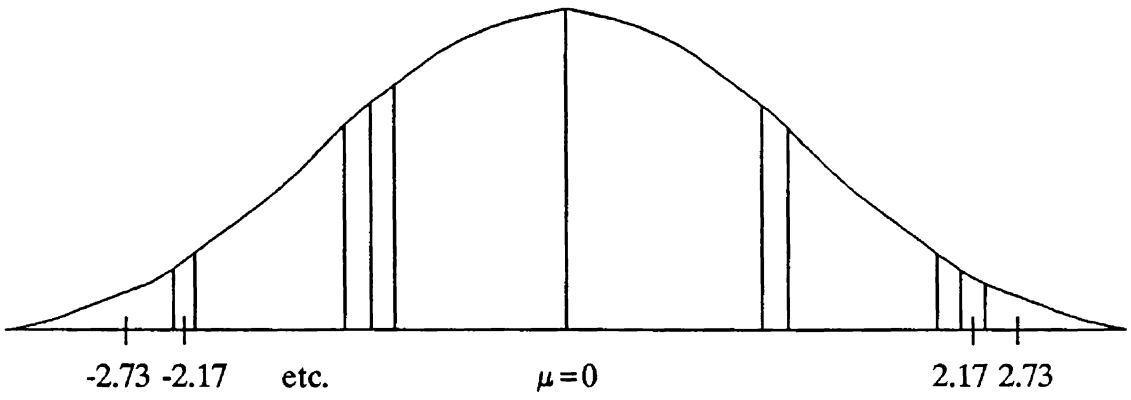
GENERATING RANDOM STANDARD NORMAL VARIABLES

Procedures used in simulating records usually utilize pseudo-random, standard normal variables. There are many ways of obtaining such variables. Nearly all, if not all, utilize pseudo-random numbers as a first step in the process.

The method described here starts with a random number. Each possible random number has a corresponding standard normal deviate to be multiplied by constants as described in Chapter 40. Random numbers can be obtained in many ways, usually with a computer routine that generates random numbers from a uniform distribution (each number within the limits of the uniform distribution is equally likely). Not all computer routines, however, are equally successful in achieving randomness in the sequence of numbers.

For practice work a table of two-digit random numbers from some source such as the table at the end of this chapter can be used which was generated by computer. A point is picked in the table by some random process. Then each succeeding two digit pair will be a random number. With this table each of 100 possible two-digit numbers from 00 to 99 is equally likely. The accompanying table gives the corresponding random normal deviates for random numbers from 00 to 99. Note that the variance of the random values is one and the mean is zero.

The table is derived by dividing the area under the normal curve into 100 equal parts symmetrical around the mean.



The segments closer to the mean will be narrower than those on the end points.

The midpoint of each segment is taken to be the value representing that area. The extreme values may be fudged somewhat to give a standard deviation of one.

Each random number corresponds to one of the segments, each of which is equally likely and results in simulation of drawing values from a random normal distribution with mean, zero, and variance, one.

To simulate distributions with mean μ and variance σ^2, the random values are modified by multiplying by σ and by adding μ.

Many computer packages are available to simulate both random numbers and random normal deviations. For simulations that are scientifically or economically important the properties of the method used in the simulation should be investigated carefully; means, variance, distribution and correlations among consecutive deviations.

TABLE FOR SIMULATING RECORDS

Random Number (00 to 99)		Standard Random	
Positive Values	Negative Values	Normal Value	
99	00	2.73	
98	01	2.17	(- if random number
97	02	1.96	between 00 and 49)
96	03	1.81	
95	04	1.70	
94	05	1.56	
93	06	1.51	
92	07	1.44	
91	08	1.37	
90	09	1.31	
89	10	1.25	
88	11	1.20	
87	12	1.15	
86	13	1.10	
85	14	1.04	
84	15	1.02	
83	16	.97	
82	17	.93	
81	18	.90	
80	19	.86	
79	20	.82	
78	21	.79	
77	22	.76	
76	23	.72	
75	24	.69	
74	25	.66	
73	26	.63	
72	27	.60	
71	28	.57	
70	29	.54	
69	30	.51	
68	31	.48	
67	32	.45	
66	33	.43	
65	34	.40	
64	35	.37	
63	36	.35	
62	37	.32	
61	38	.29	
60	39	.27	
59	40	.24	
58	41	.21	
57	42	.19	
56	43	.16	
55	44	.14	
54	45	.11	
53	46	.09	
52	47	.06	
51	48	.04	
50	49	.01	

TWO DIGIT RANDOM NUMBERS

61	16	52	80	08	91	35	16	34	56	27	83	18	98	89	61	96	25	50	73	14	18	14	03	48	79	45	01	88	33
83	28	65	29	46	92	49	67	87	84	18	22	06	68	39	90	73	25	52	80	15	57	27	11	62	17	54	86	14	30
81	43	25	33	22	69	44	25	12	96	43	12	09	56	88	11	40	24	73	26	06	61	54	27	68	02	13	25	43	33
65	90	42	46	15	65	19	34	33	72	89	53	63	46	35	87	31	70	29	66	79	17	14	08	73	04	01	53	79	01
43	17	35	96	84	99	83	72	72	37	88	99	53	62	83	59	83	30	86	23	52	76	92	35	19	51	97	37	23	39
67	09	78	98	24	92	10	04	73	08	76	53	06	06	21	74	75	57	08	56	06	38	78	49	74	56	26	80	73	25
44	35	68	94	55	79	80	26	55	58	11	22	33	04	70	66	27	34	07	95	02	82	43	41	53	32	27	48	27	10
59	00	16	42	06	26	39	38	68	61	61	12	37	61	33	98	21	21	69	38	82	94	35	96	69	75	81	64	91	53
97	39	01	36	53	84	44	15	50	70	26	67	78	55	33	42	48	05	94	37	49	41	70	28	51	50	97	45	04	80
11	52	76	13	00	40	62	05	12	94	37	57	06	30	24	61	42	29	60	99	47	95	23	21	85	59	76	56	94	46
53	24	37	14	84	08	44	37	86	05	20	72	02	97	94	43	41	80	63	37	27	82	06	15	33	64	84	72	78	83
90	30	83	30	85	22	99	54	48	56	98	79	98	65	90	43	52	79	76	22	13	24	25	71	34	27	38	35	47	33
30	09	33	98	02	46	87	75	38	19	92	71	35	83	73	88	86	31	74	37	97	78	70	34	67	97	19	26	76	32
99	69	82	51	23	61	72	95	06	55	65	73	05	32	63	97	99	35	79	94	03	57	39	57	11	73	80	79	93	74
94	42	77	93	61	51	92	06	97	19	07	88	80	13	41	06	45	91	63	78	16	91	59	89	76	12	04	54	55	95
72	91	80	49	94	09	14	61	20	93	19	62	37	79	52	92	29	15	94	93	68	82	87	41	25	15	81	98	49	80
95	16	07	96	54	81	43	32	71	59	81	45	17	02	47	42	89	98	73	98	28	33	54	36	02	85	23	97	11	77
57	92	62	53	08	83	85	11	77	92	84	00	90	46	10	96	91	08	07	86	56	51	84	62	77	36	27	97	76	97
21	62	71	33	74	63	33	71	36	99	10	66	15	73	73	32	46	09	37	49	19	79	74	01	66	31	14	63	07	35
62	55	27	77	59	27	83	03	37	67	27	07	54	02	38	51	76	62	39	62	24	05	63	78	99	24	20	20	15	23
02	55	59	56	89	24	77	82	86	22	98	28	24	67	83	82	16	94	94	13	55	76	08	34	89	77	81	94	16	14
56	41	11	70	11	51	94	11	18	43	83	29	20	80	89	66	55	46	68	06	58	42	10	11	02	59	82	07	67	83
00	52	68	66	08	00	07	47	38	19	74	14	69	40	17	57	25	98	55	42	07	69	25	07	71	00	64	79	75	06
46	73	87	28	51	61	92	82	85	27	91	31	03	70	54	76	99	03	88	31	34	09	65	32	30	90	45	26	62	12
08	67	35	51	15	42	57	66	42	61	22	81	74	00	76	37	53	19	83	72	59	13	27	63	03	66	81	65	86	87
04	07	02	78	63	62	62	69	43	32	58	58	82	34	77	07	94	06	64	69	24	05	68	40	27	55	86	17	89	53
93	18	34	39	70	07	74	04	74	05	63	79	26	41	19	17	25	08	18	38	60	20	86	69	42	52	64	97	00	26
01	89	36	74	10	98	82	67	55	79	48	18	38	67	54	24	14	12	44	01	76	37	16	73	36	74	17	43	75	75
26	08	64	78	60	71	35	87	54	94	54	18	09	50	36	93	40	74	99	47	99	03	58	90	17	63	39	33	77	21
55	09	72	89	65	62	87	62	98	28	52	56	16	62	10	13	15	67	91	28	56	13	56	52	54	94	12	04	41	02
25	53	35	12	40	31	43	24	30	40	07	65	53	42	81	80	70	46	32	82	14	40	08	95	50	18	50	25	68	88
88	22	25	15	99	30	51	36	10	07	09	62	36	85	74	69	62	09	05	54	00	97	79	36	45	87	64	57	92	50
18	48	65	26	81	06	05	26	42	13	47	59	40	12	32	45	13	18	24	55	74	61	19	75	22	35	93	47	07	37
36	38	43	68	36	07	04	94	02	63	70	28	40	17	46	79	41	68	94	47	28	26	08	35	15	35	71	18	83	17
82	85	24	00	23	16	60	43	27	29	42	09	69	40	75	40	72	71	36	79	91	21	81	89	11	38	14	04	98	48