NOTES ON THE THEORY AND APPLICATION OF SELECTION PRINCIPLES FOR THE GENETIC IMPROVEMENT OF ANIMALS

DALE VAN VLECK
DEPARTMENT OF ANIMAL SCIENCE
CORNELL UNIVERSITY
ITHACA, N. Y.

NOTES ON THE THEORY AND APPLICATION OF SELECTION PRINCIPLES FOR THE GENETIC IMPROVEMENT OF ANIMALS

Dale Van Vleck

Department of Animal Science
Cornell University
Ithaca, New York 14853

(Second Printing, 1976)
(Third Printing with Revisions, 1979)
(Fourth Printing, 1981)
(Fifth Printing with Revisions, 1983)
(Sixth Printing, 1985)
(Seventh Printing with Revisions, 1988)

PREFACE

As you will soon realize, this material is simply a compilation of mimeographed summary material which was prepared to take the place of a textbook for an upper-class and graduate student course in Animal Breeding. Consequently, these summaries are meant to supplement lectures, not replace them. The coverage is also of variable depth; the introductory summaries are quite sketchy while the summaries covering the selection index are more complete than can be found elsewhere. Most of the ideas expressed concerning the selection index have been due, in chronological order, to Sewall Wright, Jay Lush, and Charles Henderson, truly intellectual and practically-oriented giants in the development of modern animal breeding. In common with most written material, these pages will contain their share of errors--both typographical and of substance. Hopefully, neither category of error will be frequent enough to cause much harm. Finally, a word about studying this type of material. Reading some of this jargon will at first have little meaning to many people unfamiliar with the terms and symbols. Working simple problems with pencil and paper is for most people the only way to gain an understanding of the concepts. Laboratory exercises as well as the sample exam problems will aid in this endeavor. And, do not be afraid to test your solutions by a time-tried maxim--the answers should make sense.

Good luck.

Ithaca, New York

July, 1974

New topics which have been added in this revision include the selection problem with categorical traits and the effect of the fetus on performance of the dam. A major addition is an introduction to best linear unbiased prediction using mixed model equations. After working with expected values for some 12 weeks in studying the selection index, least squares equations are easy to write. Modification of these to obtain mixed model equations and, thus, best linear unbiased predictions is also relatively easy. The use of mixed model equations as pioneered by Henderson provide the statistical foundation for modern animal breeding which must jointly estimate adjustments for fixed effects and predict random effects such as transmitting abilities of dairy sires. This introduction has been tried for three to four lectures in three different years with variable success. An understanding of the ideas, however, is essential for professional animal breeders.

January, 1979

CONTENTS

SELECTED BIBLIOGRAPHY OF POSSIBLE INTEREST

Auerbach, C. *The Science of Genetics*. Paper. Harper and Row, New York. 1961.

Becker, W. A. *Manual of Procedures in Quantitative Genetics*. Mimeo. Bookstore, Washington State University, Pullman. 1967.

Bogart, R. *Improvement of Livestock*. The Macmillan Company, New York. 1959.

Brewbaker, J. L. *Agricultural Genetics*. Paper. Prentice-Hall, Inc. Englewood Cliffs, New Jersey. 1964.

Collection. *Papers on Quantitative Genetics and Related Topics*. Reproductions. Department of Genetics, North Carolina State College, Raleigh.

Crow, J. F. *Basic Concepts in Population, Quantitative and Evolutionary Genetics*. W. H. Freeman and Company, New York. 1986.

Falconer, D. S. *Introduction to Quantitative Genetics*. 2nd ed. Longman, New York. 1981.

Hutt, F. B. *Animal Genetics*. The Ronald Press Company, New York. 1964.

Hutt, F. B. *Genetic Resistance to Disease in Domestic Animals*. Cornell University Press. Ithaca, New York. 1958.

Johansson, I. *Genetic Aspects of Dairy Cattle Breeding*. University of Illinois Press, Urbana. 1961.

Johansson, I., and J. Rendel. *Genetics and Animal Breeding*. W. H. Freeman and Company, San Francisco. 1968.

Lasley, J. F. *Genetics of Livestock Improvement*. 3rd ed. Prentice-Hall, Inc., Englewood Cliffs, New Jersey. 1978.

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

Lush, J. L. *Animal Breeding Plans*. 3rd ed. Iowa State College Press, Ames. 1945.

Mather, W. B. *Principles of Quantitative Genetics*. Burgess Publishing Company, Minneapolis, Minnesota. 1964.

Pirchner, Franz. *Population Genetics in Animal Breeding*. W. H. Freeman and Company, San Francisco. 1964.

Van Vleck, L. D., E. J. Pollak, and E. A. B. Oltenacu. *Genetics for the Animal Sciences*. W. H. Freeman and Company, New York. 1987.

Warwick, E. J., and J. E. Legates. 7th ed. *Breeding and Improvement of Farm Animals*. McGraw-Hill Book Company, Inc., New York. 1979.

Summary I

QUANTIFYING THE SIMPLE MENDELIAN MODEL

The usual genetic model is

$$\text{Phenotype} = \text{Genotype} + \text{Environment},$$
$$P = G + E.$$

The simplest Mendelian model has E = 0 and only three possible genotypes 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, which has 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

We can assign arbitrary values to the genotypes as shown. The value d is for the dominance deviation of the value of the heterozygote from the average value of the homozygotes. There are several possible kinds of dominance depending on the size of d: if d = 0, we say that there is no dominance or that there is lack of dominance or that we have an additive effects model; if d = (u-v)/2, we say that there is complete dominance; and if $|d| > (u-v)/2$, we say that there is overdominance.

Population mean

The definition of the population mean or average, μ (mu), is

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

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

The mean for the simplest Mendelian model is, if there is no selection,

$$\mu = v + p(u-v), \text{ if } d = 0; \text{ and}$$
$$\mu = v + p(u-v) + 2pqd, \text{ if } d \neq 0.$$

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 to zero the derivative of μ with respect to p.

Population variance

The definition of the population variance, σ^2 (sigma squared) is

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

If $\Sigma f_i = 1$, then

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

The variance is a standard method of describing variability which is recognized by most, if not all, research workers. Note that the variance cannot be negative since it is an average of squared deviations from the mean. The variance will be more completely discussed in Summary V.

Number of genotypes

If there are n alleles at a locus, the number of possible genotypes, N, is $N = n(n+1)/2$. Note that the number of phenotypes cannot be greater than the number of genotypes and may be less.

Breeding value under simple Mendelian model

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 means of the three parental genotypes are

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

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

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

Note: (1) that $\mu_{Aa} = (\mu_{AA} + \mu_{aa})/2$ for any values of p and d, (2) that breeding value (progeny mean) of the parental genotype depends on gene frequency (even if $u > v$, μ_{AA} may be less than μ_{aa} when p is small), and (3) that the progeny frequencies are from mating a particular parent type at random to the rest of the population. For example, Aa × population gives from the gametic arrays the progeny frequencies,

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

$$\downarrow$$

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

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	$\frac{\Sigma X_i}{N} = \hat{\mu}_x = \bar{x}$	$\frac{\Sigma(X_i-\hat{\mu}_x)^2}{N-1}$	$\frac{\Sigma X_i^2-N\hat{\mu}_x^2}{N-1}$; $\frac{\Sigma X_i^2-\frac{(\Sigma X)^2}{N}}{N-1}$
Frequency data:			
Population	$\frac{\Sigma f_i y_i}{\Sigma f_i} = \mu_y$	$\frac{\Sigma f_i(y_i-\mu)^2}{\Sigma f_i}$	$\frac{\Sigma f_i y_i^2-(\Sigma f_i)\mu^2}{\Sigma f_i}$; 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 $\frac{N}{N-1}$

Summary II

SUMMARY OF MATERIAL ON POPULATION GENETICS

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 p^2 for AA, 2pq for Aa, and q^2 for aa. Note that $p + q = 1$. Hence, $q = 1 - p$, etc.

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

homozygotes	A_iA_i	with frequencies	p_i^2 for $i = 1, \ldots, n$;
heterozygotes	A_iA_j	with frequencies	$2p_ip_j$ for all $i \neq j$.

Note that $\sum_{i=1}^{n} p_i = 1$.

Estimation of gene frequencies

Generally,

$$\text{frequency of some allele} = \frac{\text{No. of that allele}}{\text{Total no. of genes at that locus}} .$$

However, the problem of estimation is illustrated in the following special cases.

(1) Dominance

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

$$q = \sqrt{\text{no. recessive types/total no. of animals}} ,$$

and $p = 1 - q$.

In the case of multiple alleles with complete dominance, we have to estimate the frequency of the most recessive allele first. For example, suppose A_1 dominant to A_2 and A_3, and A_2 dominant to A_3.

Genotypes	Phenotypes	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, $p_3 = \sqrt{\text{no. } A_3 \text{ type/total no.}}$. Plug the estimate of p_3 into the next equation, $p_2^2 + 2p_2p_3 =$ no. A_2 type/total no., and solve for p_2. Substitute the estimates of p_3 and p_2 into the remaining equation, $p_1^2 + 2p_1p_2 + 2p_1p_3 =$ no. A_1 type/total no., and solve for p_1, or find p_1 by difference since $p_1 + p_2 + p_3 = 1$, or $p_1 = 1 - p_2 - p_3$.

(2) Incomplete dominance

In this case the heterozygotes are distinguishable from the homozygotes and the gene frequencies can be found from the general formula whether or not the population is randomly mating. For example, with 3 alleles,

$$p_1 = \frac{\text{No. } A_1 \text{ genes}}{\text{Total no. genes}} .$$

Each A_1A_1 genotype contributes $2A_1$ genes; each A_1A_2 genotype contributes $1A_1$ gene; and each A_1A_3 genotype contributes $1A_1$ gene. Then,

$$p_1 = \frac{2(\text{no. of } A_1A_1 \text{ animals}) + \text{no. of } A_1A_2 + \text{no. of } A_1A_3}{2(\text{total no. of animals})},$$

and p_2 and 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 relations) equal the products of the single locus frequencies. For example, if the frequencies of A_1 and A_2 are p_1 and p_2, respectively, and the frequencies of B_1 and B_2 are r_1 and r_2, respectively, then the frequencies of the composite (two-locus) genotypes will be

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

The extension to more than 2 alleles per locus follows the same pattern.

Effect of selection on gene frequencies

Selection may change the frequency of a certain gene in a population. Gene frequency after selection (among the survivors and in the next generation) depends on the fitness of the genotypes and the gene frequencies in the

current generation. Fitness of a genotype is defined as the proportion of the genotype which reproduces relative to the other genotypes; s will be the fraction of AA genotypes which do not reproduce, r the fraction of Aa genotypes, and t the fraction of aa genotypes, where $1 \geq s, r$ and $t \geq 0$. In general, the frequency of gene a after selection is

$$q_1 = \frac{\text{no. of a genes among survivors}}{2(\text{no. of survivors})} .$$

The change in gene frequency from one generation to another is the difference in gene frequencies between the generations, i.e.,

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

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

However, in the following special cases, some simplification may be made.

(1) No homozygous recessive individuals reproduce (zero fitness for aa type)

The composition of the initial generation (n = 0) is

Genotype	Frequency	Fitness	Relative Freq. 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,

$$q_1 = \frac{\text{no. of a genes in survivors}}{\text{total no. genes in survivors}} ,$$

$$q_1 = \frac{2pq(\text{no. of animals})}{2(p^2 + 2pq)(\text{no. of animals})} = \frac{q}{1+q}$$

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 gene frequency of the recessive gene.

Corollary: The number of generations, n, required to go from a gene frequency of q to one of q_t is given by $n = (1/q_t) - (1/q)$.

(2) Selection in favor of heterozygotes

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

Genotype	Frequency	Fitness	Relative Freq. 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 gene frequency, q_1, gives

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

The change in gene 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 gene frequency from the $(n-1)^{\underline{st}}$ generation to the $n^{\underline{th}}$ generation and the population will be at equilibrium. This will be true when $sp - tq = 0$. Thus, equilibrium gene frequency will be reached when $p = t/(s+t)$ and $q = s/(s+t)$.

(3) Partial selection against homozygous recessives

The composition of the initial generation before and after selection is

Genotype	Frequency	Fitness	Relative Freq. 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 gene 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

The composition of the initial generation before and after selection is

Genotype	Frequency	Fitness	Relative Freq. 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 gene frequency, q_1, in the survivors is

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

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

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

Detection of carriers of recessive genes

The confidence of detection depends on the probability of obtaining at least one affected offspring in n offspring if the suspected carrier is actually a carrier. This is one minus the probability of obtaining all normal offspring in n offspring.

The general testing procedure is to mate a suspected carrier to a group of females which produce a fraction, P, A genes and a fraction, Q, a genes. Then, if the suspect is really a carrier, the probability that all n offspring are normal is $[1 - (Q/2)]^n$ and the confidence of detection is $1 - [1 - (Q/2)]^n$.

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 Q = 1 and that $1 - (Q/2) = 1/2$.

(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 Q = 1/2 and that $1 - (Q/2) = 3/4$.

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

The "confidence" of detecting him as a carrier is $1 - (7/8)^n$.

Note that Q = 1/4 and that $1 - (Q/2) = 7/8$. This is under the assumption that the dams were all homozygous for the normal allele.

(4) Mating a carrier male at random in a population where the frequency of the recessive gene is q in the previous generation

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

Note that $Q = q/(1+q)$ since none of the homozygous recessive females will be mated.

Note well that even one affected offspring marks a suspected carrier as a carrier. All normal offspring will never completely rule out the possibility

that a male is a carrier although the probability of detection may be quite high.

Table 1 shows the confidence of detection of carrier males for testing systems 1, 2, 3, and 4. A further discussion of method 4 follows.

AI and undesirable recessives

Method 4 of the preceding section can be used to lower 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 calves. This should yield 50 or so production-tested daughters. At the same time, the 200 calves will provide an excellent test of whether the bull is a carrier of any undesirable recessive gene.

What does this mean in terms of numbers of affected calves? We can compare an AI program which gets 200 tested calves with what would happen without AI. The effect of AI testing with 200 calves versus no testing is shown in the following table.

NUMBER OF AFFECTED OFFSPRING WITH NO PROGENY TESTING AND WITH TESTING WITH 200 PROGENY

	Before Testing		No. of affected calves per million calves born in generation 1	2	3	4	5	...	10
No testing	250,000	both 25%	111,111	62,500	40,000	27,778	20,408	...	6,944
AI test (with 200 calves)	250,000	both 25%	0	0	2	28	55	...	9
No testing	40,000	4%	27,778	20,408	15,625	12,346	10,000	...	4,444
AI test	40,000	4%	0	2	28	55	43	...	7
No testing	10,000	1% freq. of red in Holstein	8,264	6,944	5,917	5,102	4,444	...	2,500
AI test	10,000	1% freq. of red in Holstein	1	23	55	46	29	...	6
No testing	2,500	.25%	2,268	2,066	1,890	1,736	1,600	...	1,111
AI test	2,500	.25%	20	54	47	30	20	...	4
No testing	100	.01%	98	96	94	92	91	...	83
AI test	100	.01%	37	23	16	11	8	...	3

Table 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

Some Hereford breeders were at one time getting up to one-fourth dwarf calves. Dwarfism is caused by a recessive gene. With an AI testing program, that number could have been reduced to essentially none in a generation. The beef industry had not developed AI so they were limited to pedigree analysis to solve the problem. Such a system was better than no testing, but wouldn't the solution have been easy with AI?

What can AI do for the breeders of black and white Holsteins? The third line in the table shows that they can almost stop worrying about red and white calves--if their AI stud is obtaining at least 200 calves from each young bull before putting him into heavy service.

Even when the undesirable character appears very rarely such as 100 times in a million or once in every ten thousand calves, the AI testing program is much superior to natural selection. A few generations of testing will very much reduce the number of affected calves even though the reduction is not as striking as when the numbers affected are originally greater.

Most undesirables probably occur in such low frequencies. Actually, we do not worry too much about having one affected calf in ten thousand, but we really do not want as many or more than one hundred in ten thousand. We can see that an AI testing program for young sires will rapidly reduce the number of affected calves to less than one in ten thousand even if the character is very frequent before testing begins.

The reason why AI testing of young sires yields such good results is that if the frequency of affected calves is high, practically no carrier young sires will escape detection. Thus, only normal sires will be used heavily. There is no way of getting affected calves if the sire has two normal genes.

If the number of affected calves is low, carrier bulls will not be detected as well, but there will not be many carrier cows to produce affected calves so again we will not get many affected calves. Even if undetected carrier bulls do

spread the undesirable gene, almost all carriers in the next generation of bulls will be detected and the number of affected calves will be reduced still further.

With all the expressed fears that AI may sabotage a population by spreading an undesirable gene 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.

The effect of testing bulls in AI on the frequency of recessive genes

As the frequency of a recessive gene 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 gene frequencies for several generations. Males will be progeny-tested on n females. All males and females which are homozygous recessive will be culled. Heterozygotes have the same fitness as the "normal" homozygotes. Let p_j = frequency of the normal allele, A, in males surviving selection, q_j = frequency of the other allele, a, in males surviving selection, P_j = frequency of A in females surviving selection, 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 $(p_jA + q_ja)(P_jA + Q_ja)$. 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		$p_jP_j+p_jQ_j+q_jP_j$

$\alpha_i(i=j+1)$ 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. The results for various combinations of n and initial gene frequencies are shown in Tables 2, 3, and 4 and Figure 1.

Table 2. Number of affected offspring per million in the population after one generation of testing young sires.

Number of test offspring	Initial Gene Frequency .5	.2	.1	.05	.02	.01	.001
0	111,111	27,778	8,264	2,268	384	98	1.00
10	40,691	14,432	5,567	1,819	350	93	.99
25	3,422	4,478	2,952	1,297	303	87	.99
50	37	534	966	728	239	77	.97
100	0	7	96	223	147	60	.95
200	0	0	1	20	55	37	.90
500	0	0	0	0	3	8	.78
1000	0	0	0	0	0	1	.61

Table 3. Number of generations of testing young sires to reduce frequency of affected offspring in the population to various levels.

Frequency of affected offspring	Initial Gene Frequency					
	.5	.2	.1	.05	.02	.01
			No Test Offspring			
.1	2	1	1	1	1	1
.01	8	5	1	1	1	1
.001	30	27	22	12	1	1
.0001	>50	>50	>50	>50	50	1
.00001	>50	>50	>50	>50	>50	>50
			10 Test Offspring			
.1	1	1	1	1	1	1
.01	3	2	1	1	1	1
.001	10	9	7	4	1	1
.0001	30	29	27	24	15	1
.00001	>50	>50	>50	>50	>50	>50
			25 Test Offspring			
.01	1	1	1	1	1	1
.001	6	5	4	2	1	1
.0001	16	15	14	12	8	1
.00001	46	45	44	42	38	31
.000001	>50	>50	>50	>50	>50	>50
			50 Test Offspring			
.001	1	1	1	1	1	1
.0001	10	9	8	7	4	1
.00001	26	25	24	23	20	17
.000001	>50	>50	>50	>50	>50	>50
			100 Test Offspring			
.001	1	1	1	1	1	1
.0001	7	6	5	4	2	1
.00001	15	14	13	13	11	9
.000001	42	41	40	39	37	35
			200 Test Offspring			
.001	1	1	1	1	1	1
.00001	10	9	8	7	6	5
.000001	24	23	22	21	20	18
			500 Test Offspring			
.00001	1	1	1	1	1	1
.000001	13	12	11	10	9	8
			1000 Test Offspring			
.00001	1	1	1	1	1	1
.000001	10	9	8	7	6	5

Table 4. Number of affected offspring per million in the population after 35 generations of testing young sires (number of tested offspring equal 200 per sire).

Initial Gene Frequency	Generation 1	2	3	4	5	6	7	8	9	10	15	20	25	30	35
.5	.0	.0	1.7	28.0	55.4	43.4	27.7	18.1	12.5	9.1	3.0	1.4	.8	.6	.4
.2	.0	1.7	28.0	55.4	43.4	27.7	18.1	12.5	9.1	6.9	2.5	1.3	.8	.5	.4
.1	.9	22.9	54.6	45.7	29.3	19.0	13.1	9.4	7.1	5.6	2.2	1.2	.7	.5	.4
.05	20.2	53.8	46.9	30.2	19.5	13.4	9.6	7.2	5.6	4.5	1.9	1.0	.7	.4	.3
.02	55.5	41.3	26.2	17.3	12.0	8.8	6.7	5.2	4.2	3.5	1.6	.9	.6	.4	.3
.01	36.8	23.4	15.6	11.0	8.2	6.3	4.9	4.0	3.3	2.8	1.4	.8	.5	.4	.3
.001	.9	.8	.8	.7	.6	.6	.5	.5	.5	.4	.3	.2	.2	.2	.1

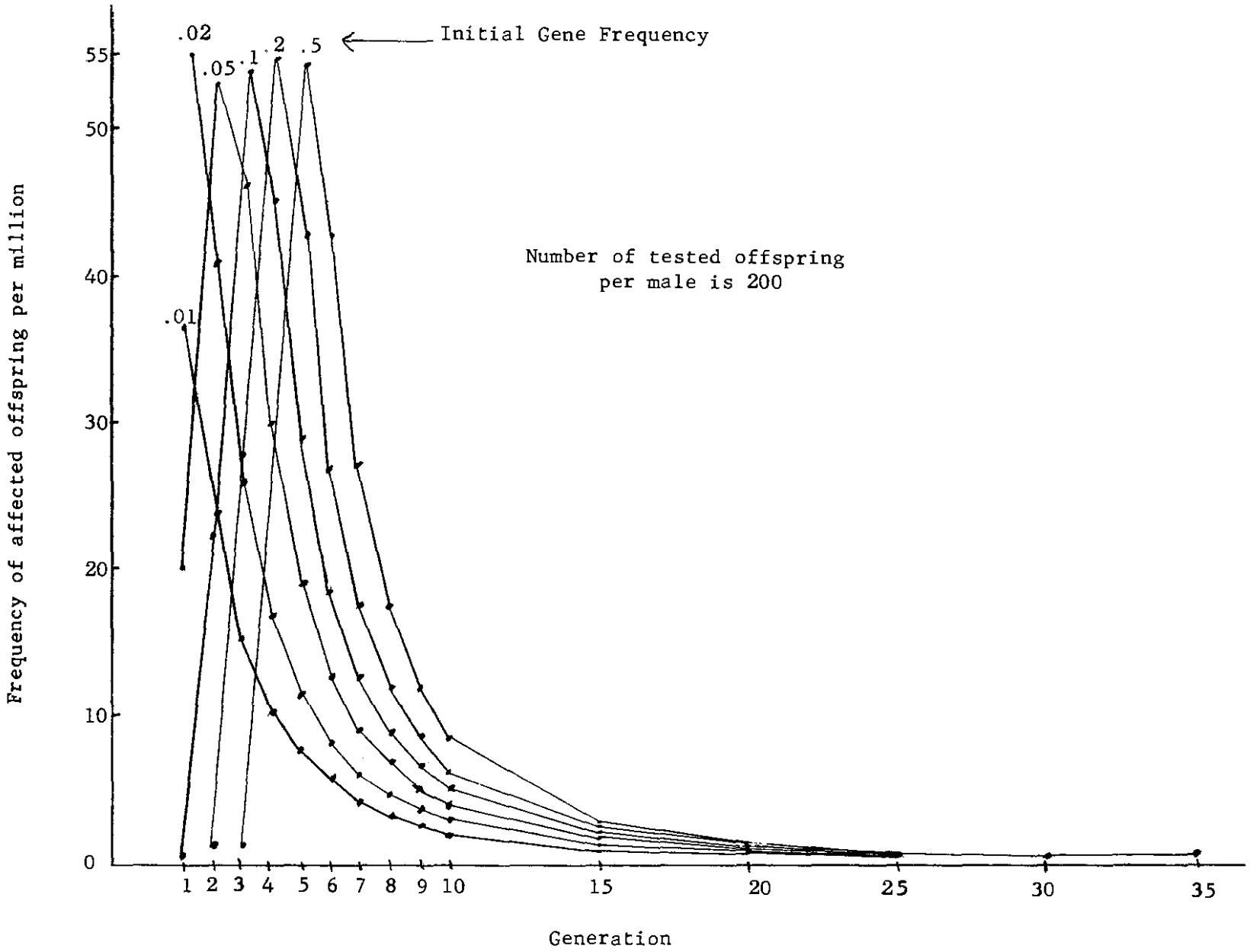


Figure 1. Frequency of affected offspring per million by generation when males are tested for heterozygosity on 200 offspring.

Summary III

REGRESSION AND CORRELATION--
DEFINING HERITABILITY FOR THE SIMPLE GENETIC MODEL

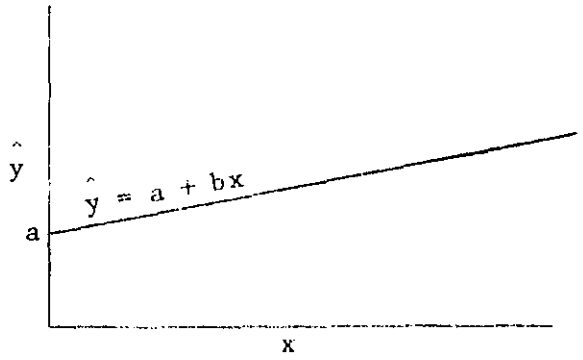
The linear regression line for predicting one variable, y, when another is known, x, is called "best" when the sum of squares of deviations from the line is minimized. The statistical equation for the regression of y on x is

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

or equivalently

$$\hat{y} = a + b_{y.x}x \ ,$$

where $a = \hat{\mu}_y - b_{y.x}\hat{\mu}_x$ and a is the intercept of the vertical axis, $b_{y.x}$ is the slope of the regression line, x is a known value, and $\hat{y}$ is the predicted value. Values of y are along the vertical axis and values of x along the horizontal axis if the relationship is plotted. Also, y will be predicted from a given value of x as $\hat{y}$.



If the regression line is estimated from a sample of data; $\hat{\mu}_y = \bar{y}$ = average value of y variable, $\hat{\mu}_x = \bar{x}$ = average value of x variable,

$$b_{y.x} = \frac{\Sigma(x-\hat{\mu}_x)(y-\hat{\mu}_y)/(n-1)}{\Sigma(x-\hat{\mu}_x)^2/(n-1)} = \frac{\left[\Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n}\right]/(n-1)}{\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right]/(n-1)}, \quad \text{where}$$

$\Sigma(x-\hat{\mu}_x)^2/(n-1)$ estimates σ_x^2 (variance of x), and $\Sigma(x-\hat{\mu}_x)(y-\hat{\mu}_y)/(n-1)$ estimates σ_{xy} (covariance of x and y). (See Summary V.)

If we have all the values in the population, $\sigma_x^2 = \Sigma(x-\mu_x)^2/n$, $\sigma_{xy} = \Sigma(x-\mu_x)(y-\mu_y)/n$, and the true regression coefficient,

$$\beta_{y.x} = \sigma_{xy}/\sigma_x^2 .$$

(Note the similarity between estimating the regression and knowing the true regression coefficient.) (Note also that $b_{y.x} = b_{x.y}$ only if $\hat{\sigma}_x^2 = \hat{\sigma}_y^2$ or if $\hat{\sigma}_{xy} = 0$.)

The population correlation coefficient is

$$r_{yx} = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2 \cdot \sigma_y^2}} = \sqrt{\beta_{y.x} \cdot \beta_{x.y}} \quad \text{with the sign of } \sigma_{xy},$$

and is estimated as

$$\hat{r}_{yx} = \frac{\hat{\sigma}_{xy}}{\sqrt{\hat{\sigma}_x^2 \cdot \hat{\sigma}_y^2}},$$

where r must be between -1 and +1 and is a standardized measure of how two variables tend to vary together.

Simple genetic model

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

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

The following rules apply:

1. Any kind of dominance will lower 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. This can also be seen by plotting progeny means against parental values.

Simple genetic model with environmental effects

Suppose for the simple genetic model we add a random environmental contribution which 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 $\sigma_p^2 = \sigma_g^2 + \sigma_e^2$. Usually we cannot separate the components of P directly. If what we want to measure is G, we may be misled by E.

Heritability defined

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

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

(In this printing, heritability will be denoted by h^2 rather than by h. As a result, there may be some misprints.)

We know that with additive gene action (the heterozygote intermediate in value between homozygotes) for the simple model that the covariance between parent and progeny is $(1/2)(\sigma^2_g)$ as shown on the previous page. This can be shown to be true even if environmental variation exists since we assume that the environmental variation is random with average value zero.

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

If there is some form of dominance, we know that this reduces the regression between progeny and parent. The covariance, however, between progeny and parent is $(1/2)(\sigma^2_{g_A})$ either with dominance or with no dominance (Appendix).

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

$$h^2 = \frac{\sigma^2_{g_A}}{\sigma^2_g + \sigma^2_e},$$

where $\sigma^2_g = \sigma^2_{g_A} + \sigma^2_{g_D}$.

Thus, twice the regression of progeny mean on parent value estimates heritability in the "narrow sense" even with dominance in the simple genetic model with random environmental effects.

Later we will see 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 Summary III

DERIVATION OF σ^2_G, $\sigma^2_{G_A}$, $\sigma^2_{G_D}$ FOR 1 LOCUS WITH 2 ALLELES

Let AA value = 1, Aa value = d, and aa value = 0. The frequency of A = p and 1 - p = q with random mating 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 G on number of "+" genes, X, to define $\sigma^2_{G_A}$ (depends on p)

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

$\sigma^2_{G_A} = r^2_{GX}\sigma^2_G$ = variance due to additive gene effects (regression of G on X).

$$= \frac{\sigma_{GX}\sigma_{GX}}{\sigma^2_X}$$

$$\mu_X = 2p$$

$$\sigma^2_X = p^2(2)^2 + 2pq - (2p)^2$$

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

$$\mu_G = p^2 + 2pqd \text{ as before.}$$

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

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

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

$$\sigma^2_{G_A} = \frac{(\sigma_{XG})^2}{\sigma^2_X} = \frac{[2pq(p+d-2pd)]^2}{2pq}$$

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

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

If $d = 1/2$, $\sigma^2_{G_A} = pq/2$. Then,

$\sigma^2_{G_D} = \sigma^2_G - \sigma^2_{G_A}$ = variance in G not accounted for by regression on X.

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

If $d = 1/2$, $\sigma^2_{G_D} = 0$.

Covariance (progeny, parent) with dominance)

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

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

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

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

which is $(1/2)(\sigma^2_{G_A})$ no matter what p or d is.

Note the values of 1, d, and 0 are scaled from general 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 $(u-v)^2$.

Summary IV

GENES IDENTICAL BY DESCENT--THE BASIS OF RELATIONSHIP

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 concept of identity by descent is a modern 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 have further developed the concept.

Two limitations are:

1. Calculations must begin at a specified base period but in all likelihood most life has originated from a small number of genes.
2. The method can only estimate how many genes are in common between two animals by descent on a probability basis.

Notation: An animal will have genes $b_i b_j$ 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 b, is identical by descent between individuals. At some arbitrary base period, tag the b genes of the common ancestor and then compute the probability that the b genes of the two individuals will be common by descent.

An example at the "b" locus:

Let the genotype of two animals be $b_i b_j$ and $b_m b_n$ where the subscript refers to the origin of the gene. We can define the probability that the genes at a locus are identical by descent

between two individuals 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, we have:

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

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

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

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

The probability that a random gene at this locus is identical in two animals is the average identical in these four comparisons, i.e., $(1/4)[P(i=m) + P(i=n) + P(j=m) + P(j=n)]$. In fact, this is the probability that a random gene from one animal and a random gene from the other will be identical by descent.

Specific example:

Mate two unrelated noninbred animals, $b_1b_2 \times b_3b_4$. The possible offspring are b_1b_3, b_1b_4, b_2b_3, b_2b_4. The fraction identical between any progeny, say b_1b_3, and any parent, say b_1b_2, is

$$P(b_1 = b_1) = 1$$
$$P(b_1 = b_3) = 0$$
$$P(b_2 = b_1) = 0$$
$$P(b_2 = b_3) = \underline{0}$$
$$\text{Average} = 1/4$$

Additive relationship

We normally think of the relationship of an individual with itself as one so the a or additive relationship between two individuals is defined as twice the fraction of genes identical by descent, and as shown in the appendix to VI, since each locus has two additive gene effects this 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 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. The following table describes most common kinds of comparisons for pairs of individuals.

POSSIBLE PROBABILITIES OF GENES IDENTICAL BY DESCENT

Comparison	Fraction Identical by Descent	"a" 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

Other examples follow:

Parent-progeny relationship

Parents, $b_1b_2 \times b_3b_4$, have 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 and the coefficient of relationship is 1/2.

Grandparent-grandprogeny relationship

Two animals, $b_1b_2 \times b_3b_4$, have 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 another animal, b_5b_6,

chosen from the population. 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 grandprogeny. The fraction of genes 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 b_1b_2 with the other 12 possible grandprogeny types. Note that 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 = 1/2, the probability of no genes in common at n loci = $(1/2)^n$ for grandparents and grandprogeny which is not a very large probability. The average identical over all loci is likely to be quite close to the probability of genes being identical by descent.

Full sib relationship

Two animals, $b_1b_2 \times b_3b_4$, have progeny b_1b_3, b_1b_4, b_2b_3, and b_2b_4. There are sixteen full sib comparisons, each having equal frequency. The values in the table are the probabilities of genes being identical for each comparison.

		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 Genotypes of 1st Full Sib with Frequencies	1/4 b_1b_3	1/2	1/4	1/4	0
	1/4 b_1b_4	1/4	1/2	0	1/4
	1/4 b_2b_3	1/4	0	1/2	1/4
	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 1 locus, and therefore the probability of no genes in common at n loci = $(1/4)^n$ for full sibs.

Half-sib relationships

Animal b_1b_2 is mated to b_3b_4, and they have progeny b_1b_3, b_1b_4, b_2b_3, and b_2b_4. Animal b_1b_2 is also mated to b_5b_6, and they have progeny b_1b_5, b_1b_6, b_2b_5, and b_5b_6. The values in the table are fraction of genes identical by descent for each half-sib comparison.

		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 Genotypes of 2nd Half-Sib with Frequencies	1/4 b_1b_5	1/4	1/4	0	0
	1/4 b_1b_6	1/4	1/4	0	0
	1/4 b_2b_5	0	0	1/4	1/4
	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 a relationship = 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

Relationship	Ave. Fraction in Common	Prob. Genotype in Common	Prob. No Genes in Common 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 2 genes are in common at one locus, i.e., for relatives with symbolic genotypes b_ib_j and b_mb_n, P(genotypes identical) = $P(b_ib_j = b_mb_n)$. Only full sibs in the above table can have a genotype at one locus identical by descent, e.g., b_1b_4 and 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):
$1/4\ b_1b_3$, $1/4\ b_1b_4$, $1/4\ b_2b_3$, $1/4\ b_2b_4$

To compute the average probability that $b_ib_j = b_mb_n$, we must find the average of all 16 comparisons as shown in the following table.

		Possible Genotypes of 2nd Full Sib with Frequencies			
		$1/4\ b_1b_3$	$1/4\ b_1b_4$	$1/4\ b_2b_3$	$1/4\ b_2b_4$
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 in common)

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

Note that only 1 of 4 comparisons are expected to have genotypes at the "B" locus in common, but 1/4 also is the average fraction of all loci with genotypes in common for one full sib with another full sib.

Dominance effects are defined as the interaction of two genes at one locus. Then dominance can contribute to likeness only between full sibs for relatives in the summary table. (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 2 genes at one locus will be identical by descent 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 2 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 genes being identical between the parents. Therefore, $F_p = (1/2)(a_{sd})$ where p, s, and d refer to the progeny, sire, and dam, respectively, and $a_{pp} = 1 + (1/2)(a_{sd})$. F_p is the inbreeding coefficient which 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

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

Also as shown in the appendix,

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

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

This equality is the basis for computing additive relationships by the tabular method.

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

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

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

$$F_A = (1/2)(a_{A_SA_D}) \; ; \; F_B = (1/2)(a_{B_SB_D}) ,$$

and an animal's additive relationship to itself is

$$a_{AA} = 1 + F_A = 1 + (1/2)(a_{A_SA_D}) \; ; \; a_{BB} = 1 + F_B = 1 + (1/2)(a_{B_SB_D}).$$

Expansion to more than one locus

The probability of a pair of nonallelic genes being alike in two individuals by descent = P(genes at first locus are identical by descent) × P(genes

at the second locus are identical by descent). Expansion to trios, etc., of nonallelic genes is obvious.

The probability of a particular combination of an allelic pair of genes (a genotype) and a nonallelic gene being identical by descent in two individuals = P(the genotypes are identical) × P(the nonallelic genes are identical).

The probability of a genotype at one locus and a genotype at another locus being common by descent in two individuals = P(first locus genotype is alike) X 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 is obvious.

To apply these principles, we need only two measures of relationship: 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 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).

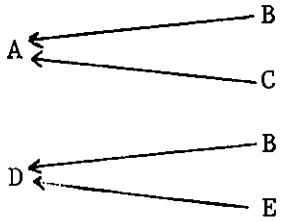
Tabular method of computing a_{ij} and d_{ij}

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

1. Determine which animals you are going to include in the table. Include all animals after the oldest 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 down 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. This 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 you know these, add the inbreeding coefficients to the diagonal.
5. Begin at the diagonal 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. It 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.

The following is an example of the a_{ij} and d_{ij} relationships for paternal half-sibs A and D.



	B	C	E	B-C A	B-E 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.,

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

Appendix to Summary IV

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

$$P(\text{random pair of genes identical}) = \alpha_{AB} = \frac{1}{4}\,[P(i=k) + P(i=\ell) + P(j=k) + P(j=\ell)]$$

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

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

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:

$$P(b_i b_j = b_k b_\ell) = P(A_S \text{ contributes } b_i \text{ to A and } B_S,\ b_{i=k} \text{ to B}) \times$$

$$P(A_D \text{ contributes } b_j \text{ to A and } B_D,\ b_{j=\ell} \text{ to B})$$

$$+ P(A_S \text{ contributes } b_i \text{ to A and } B_D,\ b_{i=k} \text{ to B}) \times$$

$$P(A_D \text{ contributes } b_j \text{ to A and } B_S,\ b_{j=\ell} \text{ to B})$$

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

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

$\alpha_{A_S B_S}$. Similarly for the other probabilities.

Thus

$$P(b_i b_j = b_k b_\ell) = \alpha_{A_S B_S} \times \alpha_{A_D B_D} + \alpha_{A_S B_D} \times \alpha_{A_D B_S}$$

$$= 1/4 \left[a_{A_S B_S} \times a_{A_D B_D} + a_{A_S B_D} \times a_{A_D B_S} \right]$$

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 $b_i b_i$ with $\alpha = 1$ and $1 - F$ of the loci are of the form $b_i b_j$ 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

$a_{AA} = 2\ \alpha_{AA} = 1 + F$.

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 $b_i b_j$ for A, $b_k b_\ell$ for S, and $b_m b_n$ 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,

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

$$= \alpha_{SD}$$

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 $b_i b_j$ for A, $b_k b_\ell$ for B_S and $b_m b_n$ for B_D.

The equally possible genotypes of B are:

B_1, $b_k b_m$;

B_2, $b_k b_n$;

B_3, $b_\ell b_m$; and

B_4, $b_\ell b_n$.

By previous definition,

$$\alpha_{AB} = \text{average of } \alpha_{AB_1},\ \alpha_{AB_2},\ \alpha_{AB_3},\ \text{and}\ \alpha_{AB_4}.$$

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

By combining and rearranging we have:

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

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

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

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

Summary V

QUANTITATIVE GENETICS: THE MEAN, STANDARD DEVIATION, AND EXPECTED VALUES

Before discussing selection for quantitative traits, a review of some basic statistics may be needed. Two important statistics for the description of continuous or quantitative data 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."

Computing the mean

If $x_i (i=1, \ldots, N)$ is the observation on the $i^{\underline{th}}$ individual, then the estimate of μ_x is $\hat{\mu}_x$ or equivalently $\bar{x}$, "x-bar";

$$\hat{\mu}_x = \sum_{i=1}^{N} x_i/N = (x_1 + x_2 + \ldots + x_N)/N = \bar{x}_N ,$$

the average of N observations. The symbol $\sum_{i=1}^{N}$ is the 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).

Variance

Although the standard deviation is more descriptive, the usual measure of variability is the variance, σ_x^2--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 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. Thus, σ_x^2 equals 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 $i^{\underline{th}}$ 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 computation 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

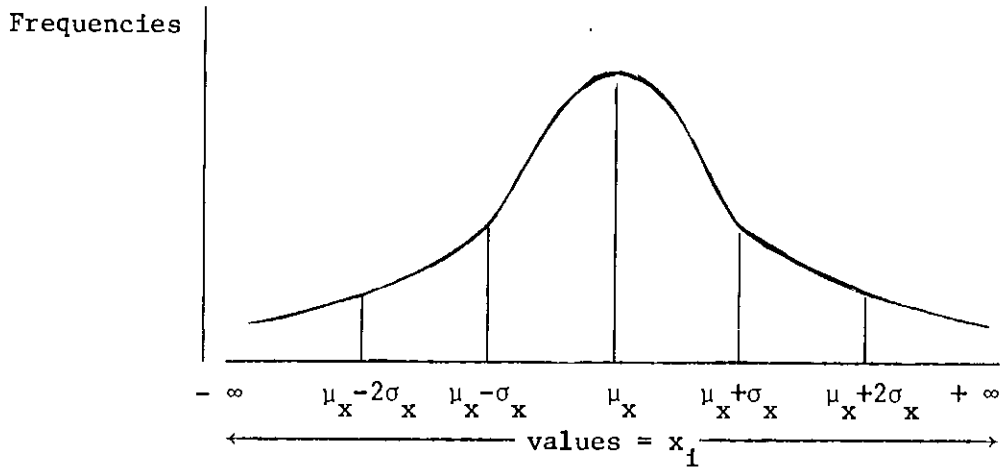
$$\hat{\sigma}_x^2 = [\Sigma x_i^2 - \frac{(\Sigma x)(\Sigma x)}{N}]/(N-1) ,$$

where N-1 is the degrees of freedom.

The division is by N-1 so that $E(\hat{\sigma}_x^2) = \sigma_x^2$, i.e., the average of estimates of σ_x^2 will be σ_x^2. Thus, the estimation procedure is said to be unbiased. Some alternative computing procedures were listed earlier on page 4. The 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 in the following fashion. Since means of observations from

other distributions approximate the normal distribution, there is sufficient justification to examine the normal distribution. 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.



μ_x is the average of all the x_i 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 $\mu_x \pm \sigma_x$ will contain 68% of the x_i; $\mu_x \pm 2\sigma_x$ will contain 96% of the x_i; etc.

The distribution of averages of N observations will have $\mu_{\bar{x}} = \mu_x$, but the variance of the averages will be $\sigma^2_{\bar{x}} = \sigma^2_x/N$.

The square root of that, $\sigma_{\bar{x}} = \sigma_{\bar{x}}/\sqrt{N}$, is often called the standard error of the mean.

Covariance

The variance thus measures 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 $i^{\underline{th}}$ individual or on the $i^{\underline{th}}$ pair of relatives. The covariance has the symbol σ_{xy} (covariance between x and y) and is defined as the average of products of deviations from the means of traits x and y;

$$\sigma_{xy} = E[(x_i-\mu_x)(y_i-\mu_y)] =$$
$$= [(x_1-\mu_x)(y_1-\mu_y) + (x_2-\mu_x)(y_2-\mu_y) + \ldots + (x_N-\mu_x)(y_N-\mu_y)]/N$$
$$= \sum_{i=1}^{N} [(x_i-\mu_x)(y_i-\mu_y)]/N \ .$$

The above procedure 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

$$\hat{\sigma}_{xy} = \Sigma[(x_i-\mu_x)(y_i-\mu_y)]/(N-1)$$
$$= [\Sigma x_i y_i - \frac{(\Sigma x_i)(\Sigma y_i)}{N}]/(N-1).$$

Note the similarity in 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.

Correlation

The correlation coefficient is a standardized measure of the relationship between two traits which allows comparisons of correlations among different traits. The possible range is -1 to +1. 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.

Expected values

After teaching selection procedures for several years and giving rules of thumb for finding variances and covariances of combinations of variables, some students suggested more facility with expected values would help in understanding the procedures. Many classes have, since then, been taught the selection index with extensive use of expected values. The use of expected values increases the powerfulness and flexibility of the selection index but at the expense of the initial frustration of many students who 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 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 find if certain definitions are remembered.

Let c = constant; x = variable from some distribution with mean μ_x and variance σ_x^2; and y = variable from some distribution with mean μ_y, variance σ_y^2, and covariance with x, 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: $E(x) = \mu_x$. The average of all possible values of variable x is its average or mean μ_x.

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

Definition 4: $E(x+y) = E(x) + E(y) = \mu_x + \mu_y$. The expectation of a sum can be taken as the sum of the expectations of the parts.

Definition 5: $E(x-\mu_x)^2 = \sigma_x^2$. By definition, the variance of a variable x, σ_x^2, is the average squared deviation of the variable from its mean. Thus, $E(x^2) = \sigma_x^2 + \mu_x^2$ which follows directly from definition 5. Expand the equation for definition 5 and take the expectations of its parts:

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

Therefore, $E(x^2) = \sigma_x^2 + \mu_x^2$. Note that $E(x^2) = \sigma_x^2$ when $\mu_x = 0$. Also, as a rule for finding the variance for a variable x, $E(x^2) = \sigma_x^2$ can be used since μ_x does not enter into the variance.

Definition 6: $E[(x-\mu_x)(y-\mu_y)] = \sigma_{xy}$. By definition, the covariance between variables x and y, σ_{xy}, is the average of the products of their deviations from their means. Thus, $E(xy) = \sigma_{xy} + \mu_x\mu_y$ which follows from definition 6. Expand the equation for definition 6 and take the expectations of its parts:

$$\sigma_{xy} = E[x-\mu_x)(y-\mu_y)]$$
$$= E(xy - \mu_x y - \mu_y x + \mu_x\mu_y)$$
$$= E(xy) - \mu_x E(y) - \mu_y E(x) + \mu_x\mu_y \text{ from (1) and (3)}$$
$$= E(xy) - \mu_x\mu_y - \mu_y\mu_x + \mu_x\mu_y$$
$$= E(xy) - \mu_x\mu_y.$$

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

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 $j^{\underline{th}}$ record in the $i^{\underline{th}}$ class, μ is a constant, A_i is a variable with $\mu_A = 0$ and variance σ_A^2, E_{ij} is a variable with $\mu_E = 0$ and variance σ_E^2, 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 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

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

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

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

The expected value of the product of observations in the same class is

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

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

The expected value of the product of observations in different classes is

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

for similar reasons as for the other expectations of the P's.

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 and any E.

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

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

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

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

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

$$Cov(P_iP_j) = E[(P_i-\mu)(P_j-\mu)] = E[(\mu + G_i + E_i - \mu)(\mu + G_j + E_j - \mu)]$$
$$= E[(G_i + E_i)(G_j + E_j)] = E(G_iG_j + G_iE_j + G_jE_i + G_jE_j)$$
$$= 0 + 0 + 0 + 0 \qquad \text{from assumptions.}$$

Practice problem

Let $X_i = \mu + E_i$, where X_i is an observation, μ is a constant, and E_i is a variable with $\mu_E = 0$ and variance σ_E^2 and the covariance between any two E's is zero. For n = 3, i = 1, 2, 3, find expectations of the following. Show all steps.

1) $E(X_i) = \mu_X =$
2) $E(X_i^2) =$
3) $E(\sum_{i=1}^{n} X_i/n) = E(\bar{X}) =$
4) $E(\Sigma X_i) =$
5) $E[\Sigma(X_i-\mu_X)] =$
6) $E[\Sigma(X_i-\mu_X)^2] =$
7) $E[\Sigma(X_i-\mu_X)^2/n] =$
8) $E[\Sigma(X_i-\bar{X})] =$
9) $E[\Sigma(X_i-\bar{X})^2] =$
10) $E[\Sigma(X_i-\bar{X})^2/(n-1)] =$

More complicated expressions are done similarly as in the example and problem.

Summary VI

GENETIC VALUES, COVARIANCES, AND EXPECTED VALUES

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 genes 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 value. 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 in addition to the normal additive gene effects of the genes. The additive by additive genetic value, G_{AA}, of an animal is the sum of all 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 and dominance 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, additive by dominance by dominance, etc. These genetic effects are defined to be independent and to have average values of zero.

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

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

If these values could be measured separately, variances for each could be computed. 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, the total genetic variance is the sum of the variances of the component genetic values:

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

A simpler but less symbolic notation for the components of genetic variance is 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 change in notation is summarized in the following table.

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

The total genetic variance can then be written as

$$\sigma_G^2 = \sum_{i+j>0} \sum \sigma_{ij}^2 = \sigma_{10}^2 + \sigma_{01}^2 + \sigma_{20}^2 + \sigma_{11}^2 + \sigma_{02}^2 + \ldots \quad .$$

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 a_{AB} which equals 2 times the probability that a random gene from A is identical by descent to a random gene from B.

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

Similarly, a_{AB}^2 is the fraction of additive by additive genetic effects in common; $a_{AB}d_{AB}$ is the fraction of additive by dominance genetic effects alike.

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

Genetic Covariances between relatives

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

Contribution to genetic covariance between individuals:

Genetic components	Contribution to covariance between individuals A and B
σ^2_{10}	$(a_{AB})^1 \ \sigma^2_{10}$
σ^2_{01}	$(d_{AB})^1 \ \sigma^2_{01}$
σ^2_{20}	$(a_{AB})^2 \ \sigma^2_{20}$
σ^2_{11}	$(a_{AB})^1(d_{AB})^1 \ \sigma^2_{11}$
σ^2_{02}	$(d_{AB})^2 \ \sigma^2_{02}$
σ^2_{12}	$(a_{AB})^1(d_{AB})^2 \ \sigma^2_{12}$
$\vdots$	$\vdots$
σ^2_{ij}	$(a_{AB})^i(d_{AB})^j \ \sigma^2_{ij}$
	($i = 0, \ldots, n$; $j = 0, \ldots, n$ with n loci, also $i + j$ must be > 0 and $i + j \leq n$)

The total genetic covariance is the sum of the parts, i.e.

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

This can be written in summation notation as

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

Note that the subscripts of the 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 values are somewhat illustrated in the coefficients in the above column for contribution to covariance between individuals. (Note that as i increases, the coefficients of the higher order genetic components of variance decrease.) 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.

Use of expected values in showing contribution to genetic covariance

In brief, the average or expected value of two variables X and Y is written E(XY). If E(X) = 0 = the average of variable X, then E(XY) = σ_{XY}. Similarly, $E(X^2) = \sigma_X^2$, $E(KX^2) = K\sigma_X^2$ where K is a constant, $E(K_X X K_Y Y) = K_X K_Y \sigma_{XY}$. These rules are given in the appendix to Summary VII and pages 45-49.

This principle will be applied to the genetic covariance between relatives X and Y for only 3 kinds of genetic effects which 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

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

The other genetic effects are due to genes from other sources and Mendelian sampling.

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

$$\sigma_{G_X G_Y} = E(G_X G_Y) .$$

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

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

(Constants can be factored outside.) Then,

$$\sigma_{G_X G_Y} = a_{XY}\sigma^2_{G_A} + 0 + 0 + 0 + 0 + 0 + 0 + 0 + d_{XY}\sigma^2_{G_D} +$$

$$0 + 0 + 0 + 0 + 0 + 0 + 0 + a^2_{XY}\sigma^2_{G_{AA}} + 0 \ .$$

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 other G_{A_Y}.

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

We know $a_{XY} = \frac{1}{2}$ and $d_{XY} = 0$.

Therefore:

$$\sigma_{G_X G_Y} = (\tfrac{1}{2})^1(0)^0\ \sigma^2_{10} + (\tfrac{1}{2})^0(0)^1\ \sigma^2_{01} + (\tfrac{1}{2})^2(0)^0\ \sigma^2_{20} +$$

$$(\tfrac{1}{2})^0(0)^2\ \sigma^2_{02} + (\tfrac{1}{2})^1(0)^1\ \sigma^2_{11} \ .$$

and so

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

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

Another example: The genetic covariance between full sibs, X and Y.

We know $a_{XY} = \frac{1}{2}$ and $d_{XY} = \frac{1}{4}$.

Then

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

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

These two examples also indicate how the components of genetic variance may be estimated. Covariance 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 (and not σ^2_{10}, σ^2_{20} since both covariances have the same expectation for those). The others usually must be assumed to be zero.

E.g., suppose Cov(Full sibs) = 50, Cov(parent-progeny) = 40. Assume $\sigma^2_{20} = \sigma^2_{02} = \sigma^2_{11} = 0$. Then

$$50 = \frac{1}{2}\sigma^2_{10} + \frac{1}{4}\sigma^2_{01}$$

$$40 = \frac{1}{2}\sigma^2_{10}$$

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

In general, for a random mating population, the additive fraction of genetic variance, σ^2_{10}, is about all we can hope to use since selection for gene combinations becomes unimportant after a few generations. Our usual goal will be to select for additive merit--the part that contributes σ^2_{10} to genetic variance.

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, $\sum\sum\sigma^2_{ij}$, 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 will be used again and again when methods of selection for additive genetic value are discussed.

Appendix to Summary VI

Reason a_{AB} describes genetic effects in common rather than $\frac{1}{2}\, a_{AB}$.

1. Additive genetic variance, σ^2_{10}

Consider one locus only.

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

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

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

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

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

$= \alpha^2\ [4P\ (\text{random genes identical by descent})]$

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

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

$= E[\alpha_i^2 + \alpha_j^2 + 2\alpha_i\alpha_j]$

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

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

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

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

2. Additive × additive genetic variance, σ^2_{20}.

Consider the minimum of two loci

and let $(\alpha\beta)_{mn}$ be the additive × additive effect of the $m^{\underline{th}}$ gene at the "a" locus and $n^{\underline{th}}$ gene at 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

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

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

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

But $\sigma^2_{20} = 4(\alpha\beta)^2$ and $\sigma^2_{20}/4 = (\alpha\beta)^2$.

Therefore $COV_{20}(AB) = \frac{\sigma^2_{20}}{4}(4a^2_{AB}) = a^2_{AB}\sigma^2_{20}$ using similar assumptions as for additive effects.

Summary VII

THE SELECTION INDEX

The basic problem in animal improvement through breeding is to choose animals which have the greatest genetic value to be parents of the next generation. The simplified model for a record (P) on an animal poses the problem;

$$P = \mu + G + E ,$$

where μ is the population mean, a constant; G is the effect on P due to the animal's complete genotype, and E is the effect of the environment on P and thus masks our evaluation of G. We have already seen that only additive genetic effects have much chance of being transmitted from one generation to another. Thus, we will usually assume that G is the additive genetic effect.

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

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

We will see later that genetic improvement per year 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, our 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 we have several animals each with records on three relatives (X_1, X_2, and X_3). We know that relatives will have genetic effects in common by descent. Thus, the record of each relative should tell us something about the genetic value of the animal we are evaluating. A logical way would be to weight each record by its relative importance, i.e., estimate of $G = I = b_1X_1 + b_2X_2 + b_3X_3$, where the b's are the appropriate weights and the X's are known records of the three relatives. I is the selection index prediction of true genetic value.

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 or 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 which is also called the accuracy of the prediction of T.
3. To maximize the probability of correctly ranking the animals, and
4. To maximize the 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 derive from work by Sewall Wright, Jay Lush, and C. R. Henderson. Henderson has 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 index is $I = b_1X_1 + b_2X_2 + \ldots + b_NX_N$ for predicting some true value, T, which will often be the additive genetic value but is not necessarily that. We want to maximize r_{TI}. Maximizing log r_{TI} is equivalent to maximizing r_{TI} but is easier to accomplish. Note that log r_{TI} = log σ_{TI} - (1/2)log σ_T^2 - (1/2)log σ_I^2.

Use of the rules for finding variances and covariances of linear functions (see Summary V) gives us σ_{TI}^2 and σ_I^2 in terms of the unknown b's and known variances and covariances. Note that σ_T^2 is a constant;

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

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

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.,

$$\frac{\partial \log r_{TI}}{\partial b_1} = \frac{\sigma_{X_1T}}{\sigma_{TI}} - \frac{b_1\sigma^2_{X_1} + b_2\sigma_{X_1X_2} + \ldots + b_N\sigma_{X_1X_N}}{\sigma_I^2} = 0$$

$$\frac{\partial \log r_{TI}}{\partial b_2} = \frac{\sigma_{X_2T}}{\sigma_{TI}} - \frac{b_1\sigma_{X_1X_2} + b_2\sigma^2_{X_2} + \ldots + b_N\sigma_{X_2X_N}}{\sigma_I^2} = 0$$

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

$$\frac{\partial \log r_{TI}}{\partial b_N} = \frac{\sigma_{X_NT}}{\sigma_{TI}} - \frac{b_1\sigma_{X_1X_N} + b_2\sigma_{X_2X_N} + \ldots + b_N\sigma^2_{X_N}}{\sigma_I^2} = 0$$

Rearrangement of these equations gives the selection index equations (except for a constant, $C = \sigma_I^2 / \sigma_{TI}$, on the right hand sides of the equations).

$$
\begin{aligned}
b_1\sigma^2_{X_1} &+ b_2\sigma_{X_1X_2} + b_3\sigma_{X_1X_3} + \cdots + b_N\sigma_{X_1X_N} &&= C\sigma_{X_1T} \\
b_1\sigma_{X_2X_1} &+ b_2\sigma^2_{X_2} + b_3\sigma_{X_2X_3} + \cdots + b_N\sigma_{X_2X_N} &&= C\sigma_{X_2T} \\
&\vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\qquad\quad \vdots &&\quad\ \vdots \\
b_1\sigma_{X_NX_1} &+ b_2\sigma_{X_NX_2} + b_3\sigma_{X_NX_3} + \cdots + b_N\sigma^2_{X_N} &&= C\sigma_{X_NT}
\end{aligned}
$$

Note:

1. The constant, $C = \sigma_I^2/\sigma_{TI}$, will not change the relative sizes of the b's or the r_{TI} so we will set C = 1, which we will see results in the same weights which minimize squared prediction error. In fact, $\sigma_I^2 = \sigma_{TI}$.
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 in row 1 and in column 1.
3. The equations are similar to multiple regression equations except the true variances and covariances are assumed known and replace the sums of squares and products used in multiple regression.
4. If $E(T-I)^2$ is minimized, the same equations would result except that no constant σ_I^2/σ_{TI} is involved on the right-hand sides of the equations.

Average squared prediction error is

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

but $\mu_T = \mu_I = 0$ or some constant that will not change differences in the I's; σ_T^2, σ_I^2, and σ_{TI} can be expressed in terms of linear functions as for the maximization of r_{TI}. Equating to zero the partial derivatives of $\sigma_T^2 + \sigma_I^2 - 2\sigma_{TI}$ with respect to b_1, b_2, ..., b_N will provide the following equations which define the b's which minimize prediction error squared;

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

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

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

These are the same equations 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.

Other properties of the selection index

The correlation between the index and true value is:

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

Note by using expected values that:

$$\sigma_{TI} = \Sigma b_i\sigma_{X_iT}$$

and that

$$r_{TI} = \sqrt{\sigma_{TI}/\sigma_T^2} \quad .$$

If the index is not the selection index, the definitional form of the correlation must be used:

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

where σ_{TI} and σ_I^2 are calculated from expected values.

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

$$\sqrt{\sigma_{TI}/\sigma_T^2}$$

because $\sigma_I^2 = \sigma_{TI}$.

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

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

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

When I is not the selection index

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

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

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

This expression corresponds to the variance 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:

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

The average of true values for animals with index value I_0 is:

$$E(T|I=I_0) = I_0 \quad .$$

Thus the selection index procedure is unbiased.

Intuitively, we would expect animals with the same index value to have different true values. In fact, the variance of true values for animals with the same index value, I_0, is:

$$\sigma^2_{T|I=I_0} = (1-r_{TI}^2)\sigma_T^2 \quad .$$

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

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.

Models for determining covariances between records

Single measurement 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 $\sigma_{G_iG_j} = a_{ij}\sigma^2_{10} + d_{ij}\sigma^2_{01} + \ldots$ as developed in Section VI.

For convenience of notation we will define the covariance between environmental effects on records of relatives i and j 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 is often a good approximation.

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

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

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 every 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 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 repeatability model and sometimes the animal model. The sum of all permanent effects of the animal can be denoted as the animal effect :

$$A_i = G_i + PE_i \quad .$$

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

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

Note that $\sigma_A^2 = r\sigma_X^2$.

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

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

under the assumption of no covariances between animal effect and temporary environmental effects and between temporary environmental effects.

How to determine the variance of an average

Let X_i be the average of N_i records:

$$X_i = \frac{X_{i1} + \ldots + X_{iN_i}}{N_i} \quad .$$

If $E(X_{ij}^2) = \sigma_X^2$ for all ij (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 have the same covariance), then from expected values:

$$\sigma_{X_i}^2 = E(X_i^2) = E\left[\left(\frac{X_{i1} + \ldots + X_{iN_i}}{N_i}\right)^2\right]$$

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

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

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 and

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

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

each contributing a record to the average and

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

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 contributing to the genetic covariance), then

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

General form of the covariance between averages

The covariance between averages is usually 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

$$X_j = \frac{X_{i1} + \ldots + X_{iN_i}}{N_i} \quad ,$$

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

$$X_j = \frac{X_{j1} + \ldots + X_{jN_j}}{N_j} \quad .$$

If $E(X_{ik}X_{j\ell})$ is the same for all k and ℓ, then taking expected values shows from

$$Cov(X_i,X_j) = E\left(\left(\frac{X_{i1} + \ldots + X_{iN_i}}{N_i}\right)\left(\frac{X_{j1} + \ldots + X_{jN_j}}{N_j}\right)\right)$$

that there are N_iN_j expected values with the same expectation in the numerator with a denominator of N_iN_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

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

Summary

1) If X_i is the average of records on relative i, the $\sigma_{X'X} = r\sigma_X^2$,

2) If X_i is the average of single records of relatives of type i, then $\sigma_{X'X} = \sigma_{G_iG_i'} + C_{ii'}\sigma_X^2$, and

3) If X_i is the average of n_i records on each of p_i relatives of type i, then also $\sigma_{X'X} = \sigma_{G_iG_i'} + C_{ii'}\sigma_X^2$.

The derivation of the variance of an average of averages can be done with expected values utilizing the 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 $X_i = \dfrac{\bar{X}_{i1} + \ldots + \bar{X}_{ip_i}}{P_i}$ where $\bar{X}_{ij}$ is the average of n_i records on animal j in relative group i. The number of animals in group i is p_i. Then

$$\sigma^2_{X_i} = E\left(\left(\frac{\bar{X}_{i1} + \ldots + \bar{X}_{iP_i}}{P_i}\right)^2\right) = E\left(\frac{(\bar{X}_{i1})^2 + \ldots + (\bar{X}_{iP_i})^2 + \text{all products}}{P_i^2}\right)$$

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

$$= \frac{\dfrac{\sigma_X^2 + (n_i-1)r\sigma_X^2}{n_i} + (p_i-1)(\sigma_{G_iG_i'} + C_{ii'}\sigma_X^2)}{P_i} \quad .$$

How to determine the σ_{X_iT}

If we know $\sigma^2_{X_i}$, $\sigma_{X_iX_j}$, and σ_{X_iT} for all i and j, we can easily set up the equations to find the appropriate weights for the index; $\sigma^2_{X_i}$ and $\sigma_{X_iX_j}$ can be estimated or derived as shown, but σ_{X_iT} is the covariance between something we can measure, X_i, and T, something we cannot measure or see. Therefore, σ_{X_iT} must be computed indirectly. If we are selecting 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 individual α which we are evaluating for G_{A_α}. σ^2_{10} is the additive genetic variance. We can see that this is the portion of the genetic covariance between relatives which is due to additive genetic effects in common between the relatives. 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 procedure 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, which we are trying to predict. The only parameters of the procedure that change when T is redefined are the right-hand sides, the σ_{X_iT}, and also σ^2_T although other parameters that depend on these will, of course, also change. Expected values and simple models can be used to find easily σ_{X_iT} and σ^2_T. The technique will be demonstrated for several definitions of T including the usual one where T is additive genetic value. To simplify expected values, all variables will be assumed to have zero means although

as stated earlier the variances will be the same in either 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 value, etc. PE_i is a permanent environmental effect on all records of animal i, and TE_i is a temporary environmental effect on a specific record of i. Then,

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

   if no genetic by environmental correlation. Thus, the right-hand sides 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 $\sigma^2_{10} = h^2\sigma^2_X$ if α is not inbred.

2. $T = A_\alpha = G_\alpha + PE_\alpha$, real producing ability (i.e., prediction of the permanent ability).

   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 $i \neq \alpha$, $\sigma_{X_iT} = E(G_i + PE_i + TE_i)(G_\alpha + PE_\alpha)] = E(G_iG_\alpha) + E(PE_iPE_\alpha)$ + others likely to be zero $= \sigma_{G_iG_\alpha} + \sigma_{PE_iPE_\alpha}$ (total genetic covariance plus permanent environmental covariance usually assumed to be zero but not necessarily so, e.g., for littermates). If all $i \neq \alpha$, the index weights will be the same as for predicting additive genetic value, but σ^2_T will be different; $\sigma^2_T = E(A^2_\alpha) = E[(G_\alpha + PE_\alpha)^2] = \sigma^2_A = \sigma^2_G + \sigma^2_{PE}$, if not inbred. Repeatability or the correlation between records on the same animal is defined as $r = (\sigma^2_G + \sigma^2_{PE})/\sigma^2_X = \sigma^2_A/\sigma^2_X$.

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

$$\sigma_{X_i T} = E[(G_i + E_i)(G_{D_\alpha})] = d_{i\alpha}\sigma^2_{01}$$
$$\sigma^2_T = E(G^2_{D_\alpha}) = \sigma^2_{01}$$

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

$$\sigma_{X_i T} = E[(G_i + E_i)(G_{A_\alpha} + G_{D_\alpha})] = a_{i\alpha}\sigma^2_{10} + d_{i\alpha}\sigma^2_{01}$$
$$\sigma^2_T = E[(G_{A_\alpha} + G_{D_\alpha})^2] = \sigma^2_{10} + \sigma^2_{01}, \text{ if not inbred.}$$

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

$$\sigma_{X_i T} = E[(G_i + E_i)(G_\alpha)] = \sigma_{G_i G_\alpha} = a_{i\alpha}\sigma^2_{10} + d_{i\alpha}\sigma^2_{01} + \ldots$$
$$\sigma^2_T = E(G_\alpha) = \sigma^2_G = \sigma^2_{10} + \sigma^2_{01} + \ldots ,$$

if not inbred. If $G_\alpha = G_A$, then everything is the same as for predicting additive genetic value.

6. $T = (1/2)G_{A_\alpha}$, the average part of additive genetic value that is transmitted to progeny--transmitting ability.

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

Thus, the index weights will only be half as large as when predicting α's additive genetic value.

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

The six definitions above show the flexibility of the selection index as long as T can be defined. In other cases, there is more difficulty in determining exactly what T is:

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

If $i = \alpha$ (animal already has a record, e.g., record P_{α_1} and want to predict from this record, record P_{α_2}),

$$\sigma_{X_i T} = E[(G_\alpha + PE_\alpha + TE_{\alpha_1})(G_\alpha + PE_\alpha + TE_{\alpha_2})]$$
$$= \sigma^2_G + \sigma^2_{PE} = r\sigma^2_X = \sigma^2_A ,$$

if not inbred. If $i \neq \alpha$,

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

The last two terms are environmental covariances which often are assumed to be zero. The first term is the total genetic covariance. These right-hand sides and index weights are the same as for predicting real producing ability if $E(TE_iTE_\alpha) = 0$, but σ_T^2 is different; $\sigma_T^2 = E(G_\alpha + PE_\alpha + TE_\alpha)^2 = \sigma_G^2 + \sigma_{PE}^2 + \sigma_{TE}^2 = \sigma_G^2 + \sigma_E^2 = \sigma_P^2 = \sigma_X^2$, the total phenotypic variance of single records.

8. T = average of records of m future half-sib progeny of some sire = $[(\Sigma G_i)/m] + [(\Sigma E_i)/m]$

Since the covariance between averages and between individual records is the same, let $P_\alpha = G_\alpha + E_\alpha$ be a representative record in T.

$$\sigma_{X_iT} = E[(G_i + E_i)(G_\alpha + E_\alpha)] = \sigma_{G_iG_\alpha} \quad (= a_{i\alpha}\sigma_{10}^2 \text{ if } G_\alpha = G_{A_\alpha}).$$

However,

$$\sigma_T^2 = E\{[\frac{\Sigma(G_i + E_i)}{m}]^2\} = \frac{\sigma_X^2 + (m-1)\sigma_{XX'}}{m}, \text{ (see Appendix)}$$

where $\sigma_{XX'}$ 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, $\sigma_{G_iG_{i'}} + \sigma_{E_iE_{i'}}$, the genetic plus environmental covariance between i and i' are both included in it.

9. T = average of records of ∞ future half-sib progeny of some sire.

σ_{X_iT} as in (8), but $\sigma_T^2 = (\sigma_X^2/m) + [(m-1)\sigma_{XX'}/m]$, and as $m \to \infty$, $\sigma_T^2 \to \sigma_{XX'} = \sigma_{G_iG_{i'}} + \sigma_{E_iE_{i'}}$. This case is often similar to predicting $(1/2)(G_{A_\alpha})$ of a sire when $G_\alpha = G_{A_\alpha}$.

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

$\sigma_{X_i T} = a_{i\alpha}\sigma^2_{10}$ as in (8); $\sigma^2_T = E\{[(\Sigma G_{A_\alpha})/m]^2\} = [\sigma^2_{10} + (m-1)a_{\alpha\alpha'}\sigma^2_{10}]/m$ since σ^2_{10} is the variance of additive genetic value and $a_{\alpha\alpha'}\sigma^2_{10}$ is the covariance between additive genetic value of α and α', a representative pair in the group. As $m \to \infty$, $\sigma^2_T \to a_{\alpha\alpha'}\sigma^2_{10}$. For noninbred half-sib progeny, $a_{\alpha\alpha'} = 1/4$ and $\sigma^2_T = (1/4)\sigma^2_{10}$ as in case 6 when predicting $(1/2)(G_{A_{sire}})$. Note that in (10), α refers to a progeny group, and in (6), α designates a particular animal 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 also should be clear and would avoid much of the confusion among animal breeders who seem to be selecting for the same thing when they actually are not.

X_i are averages of records of a relative i

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 we can assume, as often is nearly true, that the variance of first records equals the variance of second record, 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 weights is

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

The off-diagonal coefficients are the same as the covariances between a single record of one animal and a single record of another relative.

If, however, we can assume that the only reason for likeness between relatives is common additive genetic effects, then the off-diagonal coefficients are of the form $\sigma_{X_i X_j} = a_{ij}\sigma_{10}^2 = a_{ij}h^2\sigma_X^2$. If other components of genetic variance are important, this will not be true but still may be a reasonable approximation if the true covariance is unknown 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 $C_{ij}\sigma_X^2$ is the covariance between records of relatives i and j caused by common environmental effect, then the off-diagonal coefficient should be $\sigma_{X_i X_j} = (a_{ij}h^2 + C_{ij})\sigma_X^2$. The equations to find the b's can now be written (assuming all $C_{ij} = 0$) to predict G_{A_α}:

$$\begin{aligned}
&\left(\frac{1+(n_1-1)r}{n_1}\right)\sigma_X^2 b_1 + a_{12}h^2\sigma_X^2 b_2 + a_{13}h^2\sigma_X^2 b_3 + \ldots + a_{1N}h^2\sigma_X^2 b_N = a_{1\alpha}h^2\sigma_X^2 \\
&a_{12}h^2\sigma_X^2 b_1 + \left(\frac{1+(n_2-1)r}{n_2}\right)\sigma_X^2 b_2 + a_{23}h^2\sigma_X^2 b_3 + \ldots + a_{2N}h^2\sigma_X^2 b_N = a_{2\alpha}h^2\sigma_X^2 \\
&\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\
&a_{1N}h^2\sigma_X^2 b_1 + a_{2N}h^2\sigma_X^2 b_2 + a_{3N}h^2\sigma_X^2 b_3 + \ldots + \left(\frac{1+(n_N-1)r}{n_N}\right)\sigma_X^2 b_N = a_{N\alpha}h^2\sigma_X^2
\end{aligned}$$

We can see that σ_X^2 appears in each equation so that dividing each equation by σ_X^2 will not change the solutions for the b's.

Thus, the equations can be written as:

$$\begin{aligned}
&\frac{1+(n_1-1)r}{n_1} b_1 + a_{12}h^2 b_2 + a_{13}h^2 b_3 + \ldots + a_{1N}h^2 b_N = a_{1\alpha}h^2 \\
&a_{12}h^2 b_1 + \frac{1+(n_2-1)r}{n_2} b_2 + a_{23}h^2 b_3 + \ldots + a_{2N}h^2 b_N = a_{2\alpha}h^2 \\
&\vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \vdots \\
&a_{1N}h^2 b_1 + a_{2N}h^2 b_2 + a_{3N}h^2 b_3 + \ldots + \frac{1+(n_N-1)r}{n_N} b_N = a_{N\alpha}h^2
\end{aligned}$$

Only r and h^2 are necessary in order to set up the equations because the relationships can be computed and the n's will be known.

X_i is the average of n_i records on p_i equally related relatives

Suppose X_i is the average of a genetic group of animals (p_i) each with n_i records (e.g. a group of paternal half-sisters each with 2 records). Further suppose 1) each animal in the group has the same relationship, $a_{ii'}$, to all other animals in the group, and 2) each animal in one group has the same relationship to all animals in each other group, i.e., a_{ij}'s are same for all animals in groups i and j. Then the diagonal coefficients become

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

If other than additive genetic variance contributes to likeness between animals in the genetic group, the portion $(p_i - 1)\, a_{ii'}h^2$ 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 coefficient is:

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

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

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

$$\frac{1+(n_1-1)r}{n_1} + (p_1-1)a_{11}, h^2 \Big/ p_1 \; b_1 + a_{12}h^2 \; b_2 + \ldots + a_{1N}h^2 b_N = a_{1\alpha}h^2$$

$$a_{12}h^2 b_1 + \frac{1+(n_2-1)r}{n_2} + (p_2-1)a_{22}, h^2 \Big/ p_2 \; b_2 + \ldots + a_{2N}h^2 b_N = a_{2\alpha}h^2$$

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

$$a_{1N}h^2 b_1 + a_{2N}h^2 b_2 + \ldots + \frac{1+(n_N-1)r}{n_N} + (p_N-1)a_{NN}, h^2 \Big/ p_N \; b_N = a_{N\alpha}h^2$$

The assumptions which are implied in using this 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 covariances among records for each animal are $r\sigma_X^2$ for all relatives.

4) Only additive genetic variance contributes to the covariance among relatives. If this assumption is not true the terms $a_{ij}h^2$ should be modified to take into account other components of genetic variance and any environmental covariance.

5) Each animal in group $\underline{i}$ has the same number of records. If not, the group should be divided so that each new group has the same number of records.

Records from inbred animals

The variation among non-related 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. Thus, the diagonal coefficients of the equations which determine the selection index weights will be increased. For single records the increase will be $F_i h^2 \sigma^2_X$ to $(1+F_i h^2)\sigma^2_X$. For the average of records on the same animal the diagonal coefficient will be

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

since the covariance between records on the same animal will be also increased by $F_i h^2 \sigma^2_X$.

For the average of single records on each of p_i animals in group i, the diagonal coefficient will be

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

This situation, however, seems rather unlikely. For the average of n_i records on each of p_i animals in group i, the diagonal coefficient becomes

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

If animal α is inbred, r_{TI} will also have $(1+F_\alpha)h^2\sigma^2_X$ in the denominator as σ^2_T when selecting for G_{A_α}.

Computation of r_{TI}

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

$$r^2_{TI} = \Sigma b_i \sigma_{X_i T} / \sigma^2_T \quad .$$

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

Then

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

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

Computation of $\sigma_T^2|_{I=I_o}$

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

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

Additional note

Often 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 a set of equations corresponding to the p_i's and n_i's associated with that animal.

Application of the index to cases where the assumptions are true

1. One or several records per individual.

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 of the RHS's all are equal then the equations are simplified.

Repeatability in the case of equal variances and covariances is defined as $r = \frac{\sigma_{X_i X_j}}{\sigma^2_{X_i}}$, the regression of the jth record on the ith record or the correlation between the i and jth records.

If the covariances between all records and the additive genetic value of the individual are all equal, then the index becomes

$$I = bX$$

where X is the average of n records on an individual for which we want to predict additive genetic value.

The equation to find b is:

$$\left(\frac{1 + (n-1)r}{n}\right) b = h^2 \text{ for equal variances and covariances}$$

and $b = \frac{nh^2}{1+(n-1)r}$.

$$r_{TI} = \sqrt{\left(\frac{nh^2}{1+(n-1)r}\right) \times h^2\sigma^2_X/h^2\sigma^2_X} = \sqrt{\frac{nh^2}{1+(n-1)r}} \text{ since } \sigma^2_T = h^2\sigma^2_X = \sigma^2_{10}$$

$$\sigma^2_{T|I=I_o} = \left(1 - \frac{nh^2}{1+(n-1)r}\right) h^2\sigma^2_X \text{ for animals with the same number of records}$$

and the same index values.

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 the αth individual.

The index equations will be:

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

Only additive relationships and heritability are needed to set up the index equations.

3. The case where some of the related individuals have more than one record (n_i).

Now the diagonals will be $\frac{1+(n_i-1)r}{n_i} b_i$ instead of b_i. The off-diagonals and RHS's will be the same as before.

4. The case where we have the averages of single records of p_i members of groups with relationship $a_{ii'}$ with each other, all having the same relationship to α and to other groups or individuals used in the index.

Now the diagonals become

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

The off diagonals and RHS's are the same as before.

5. The case where members of the groups have more than one record (n_i).

The diagonals now become

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

The off diagonals and RHS's are the same as before. This is the general form of the diagonal coefficients because when $n_i = 1$, the diagonal is the same as for case 4 and when $p_i = 1$, the diagonal is the same as for case 3.

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

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)	$h^2/2$	$\sqrt{h^2}/2$
	(n)	$nh^2/2[1 + (n-1)r]$	$\sqrt{nh^2/[1 + (n-1)r]}/2$
Sire and dam	(1)	$h^2/2$; $h^2/2$	$.71\sqrt{h^2}$
	(n)	$nh^2/2[1 + (n-1)r]$; $nh^2/2[1 + (n-1)r]$	$.71\sqrt{nh^2/[1 + (n-1)r]}$
One grandparent		$h^2/4$	$\sqrt{h^2}/4$
Four grandparents		All $h^2/4$	$\sqrt{h^2}/2$
One great-grandparent		$h^2/8$	$\sqrt{h^2}/8$
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)		$2ph^2/[4 + (p-1)h^2]$	$\sqrt{ph^2/[4 + (p-1)h^2]}$

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)	$[h^2D - (h^2/2)^2]/[C - (3h^4/16)]$; $h^2(A-h^2)/2[C - (3h^4/16)]$	$\sqrt{b_1 + (b_2/2)}$
Dam (n) and paternal half-sibs (p)	$nh^2/2[1 + (n-1)r]$; $ph^2/[4 + (p-1)h^2]$	$\sqrt{(b_1/2) + (b_1/4)}$
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)	$mh^2/[4 + (m-1)h^2]$; $h^2[D - (h^2/16)]/2C$ $h^2(A-h^2)/8C$	$\sqrt{(b_1/4) + (b_2/2) + (b_3/8)}$

Summary VIII

SIRE EVALUATION

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

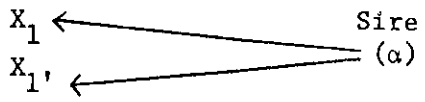
The basic problem is simple:

GIVEN: X, the average of single records of p progeny all from different dams.

ESTIMATE: Additive genetic value of sire from I = bX.

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

$$\frac{1 + (p-1)\, a_{11'} h^2}{p}\, b = a_{1\alpha} h^2.$$



In this situation, $a_{11'} = \frac{1}{4}$ and $a_{1\alpha} = \frac{1}{2}$. Thus,

$$b = \frac{\frac{1}{2} ph^2}{1+(p-1)\frac{1}{4}h^2} = \frac{2p}{p + \frac{4-h^2}{h^2}}.$$ As $p \longrightarrow \infty$, $b \longrightarrow 2$.

For $h^2 = \frac{1}{4}$, $b = \frac{2p}{p+15}$; for $h^2 = \frac{1}{2}$, $b = \frac{2p}{p+7}$; etc.

In general, $r_{TI} = \sqrt{a_{1\alpha} b} = \sqrt{\frac{1}{2}\left(\frac{2p}{p+\frac{4-h^2}{h^2}}\right)} = \sqrt{\frac{p}{p+\frac{4-h^2}{h^2}}}$.

As $p \longrightarrow \infty$, $r_{TI} \longrightarrow 1$.

Note: 1) That we don't have to set up a new equation for each sire with different p since we have solved for b in terms of p and h^2;

2) b depends on p,

3) r_{TI} depends on p, and

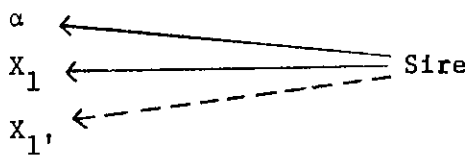
4) b can exceed 1. In genetic evaluation, the b's are usually less than 1 except for sire evaluation from progeny records.

Variations on sire evaluation

The above is the basis for estimating the additive genetic value of the sire. Some similar procedures will yield a b which is $\frac{1}{2}$ the b we found. For example, a Cornell Sire Evaluation Procedure which was called the Cornell Daughter Level Report used $\frac{p}{p+15}$ rather than $\frac{2p}{p+15}$. The U.S.D.A. procedure also uses a similar form although the heritability they use is slightly lower.

These are two reasons for using the smaller weight:

A. Rather than estimating the genetic value of the sire, we are interested in estimating the genetic value of a future daughter, α.



$a_{11'} = \frac{1}{4}$ as before, but $a_{1\alpha} = \frac{1}{4}$ rather than $\frac{1}{2}$.

Now, $b = \dfrac{p}{p + \frac{4-h^2}{h^2}}$ and as $p \longrightarrow \infty$, $b \longrightarrow 1$.

For $h^2 = \frac{1}{4}$, $b = \dfrac{p}{p+15}$; for $h^2 = \frac{1}{2}$, $b = \dfrac{p}{p+7}$; etc.

Also $r_{TI} = \sqrt{\frac{1}{4}\left(\dfrac{p}{p + \frac{4-h^2}{h^2}}\right)} = \frac{1}{2}\sqrt{\dfrac{p}{p + \frac{4-h^2}{h^2}}}$ and as $p \longrightarrow \infty$, $r_{TI} \longrightarrow \frac{1}{2}$. Note this is the accuracy of predicting the genetic value of an animal from records of p paternal half sisters.

For $h^2 = \frac{1}{4}$, $r_{TI} = \frac{1}{2}\sqrt{\dfrac{p}{p+15}}$; for $h^2 = \frac{1}{2}$, $r_{TI} = \frac{1}{2}\sqrt{\dfrac{p}{p+7}}$; etc.

B. The daughter or progeny superiority of a sire (also called transmitting ability) may be estimated. The daughter superiority is defined to be the average of an infinite number of future daughters or one-half the additive genetic value of the sire, i.e.,

$T = \frac{1}{2} G_{SIRE}$. Then $\sigma_{X_1 T} = \sigma_{X_1 \frac{1}{2}G} = \frac{1}{2} a_{1\alpha} h^2 \sigma_X^2 = \frac{1}{4} h^2 \sigma_X^2 = \frac{1}{4}\sigma_{10}^2$ since $a_{1\alpha} = \frac{1}{2}$.

The equation to find b is:

$$\frac{1 + (p-1)\frac{1}{4}h^2}{p} b = \frac{1}{2}\left(\frac{1}{2}\right) h^2$$

and $b = \dfrac{p}{p + \frac{4-h^2}{h^2}}$ as in A.

r_{TI} , however, will not be the same as in A.

Remember $r_{TI} = \sqrt{\dfrac{b\ \sigma_{X_1 T}}{\sigma_T^2}}$. Note that $b\ \sigma_{X_1 T} = \left(\dfrac{p}{p + \frac{4-h^2}{h^2}}\right) \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) \sigma_{10}^2$.

But $T = (1/2)G$. Thus, by the rules, $\sigma_T^2 = E(T^2) = E[(G/2)^2] = (1/4)E[\ G\ ^2] = (1/4)\sigma_{10}^2$ since G is additive genetic value. Thus,

$$r_{TI} = \sqrt{\frac{\dfrac{p}{p + \frac{4-h^2}{h^2}}\left(\frac{1}{4}\ \sigma_{10}^2\right)}{\left(\frac{1}{4}\ \sigma_{10}^2\right)}} = \sqrt{\frac{p}{p + \frac{4-h^2}{h^2}}}$$ which is the same as

when estimating the additive genetic value of the sire. This is reasonable since we have divided by a constant one-half.

Note well: It is important to define T exactly since what T is makes a difference in both b and 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 exists among animals in a progeny group.

Suppose that there is an environmental correlation among half-sibs

in the same environment, then the environmental covariance is $c_{11'}\sigma_X^2$.

The equation to find the b to evaluate the sire from p progeny with one record each is:

$$\frac{1 + (p-1)[a_{11'}h^2 + c_{11'}]}{p} \quad b = a_{1\alpha}h^2$$

where $a_{11'}$ is the relationship among animals in the group, $a_{11'} = \frac{1}{4}$ if half sibs.

$c_{11'}\sigma_X^2$ is the environmental covariance and

$a_{1\alpha}$ is the relationship of animals in the group to α. If α is the sire, then $a_{1\alpha} = \frac{1}{2}$.

Thus, $b = \frac{\frac{1}{2}ph^2}{1 + (p-1)(\frac{1}{4}h^2 + c_{11'})}$ and $r_{TI} = \sqrt{\frac{\frac{1}{4}ph^2}{1 + (p-1)[\frac{1}{4}h^2 + c_{11'}]}}$.

If $c_{11'} = \frac{1}{4}h^2$ and $h^2 = \frac{1}{4}$ as is approximately true for dairy production:

$b = \frac{p}{p+7}$ or $\frac{2p}{p+14}$ rather than $\frac{2p}{p+15}$ and

$$r_{TI} = \sqrt{\frac{p}{p+7}\left(\frac{1}{2}\right)} = .71\sqrt{\frac{p}{p+7}} \quad .$$

Note that in this case as $p \longrightarrow \infty$, $b \longrightarrow 1$ but $r_{TI} \longrightarrow .71$.

If, $c_{11'} \neq 0$, then as $p \longrightarrow \infty$, $r_{TI} \longrightarrow$ less than 1.

The following table compares the b's and r_{TI}'s when $h^2 = \frac{1}{4}$ with and without environmental correlation.

Environmental covariance	$c_{11'} = 0$		$c_{11'} = 1/16$	
	No environmental correlation		Environmental correlation	
p	$b = \frac{2p}{p+15}$	$r_{TI} = \sqrt{\frac{p}{p+15}}$	$b = \frac{p}{p+7}$	$r_{TI} = .71\sqrt{\frac{p}{p+7}}$
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

Note that the b's are smaller and $r_{TI} \longrightarrow .71$ for $c_{11'} = 1/16$.

This table assumes the environmental correlation is the same for all pairs of progeny. The U.S.D.A. sire evaluation procedure, however, considers the situation where only daughters in the same herd have an environmental correlation.

If there are n_i daughters in the $i^{\underline{th}}$ herd, then

$$b = \frac{2\,ph^2}{4 + (p-1)\,h^2 + \frac{4\Sigma n_i (n_i-1)\;\; c_{11'}}{p}} \quad . \quad \text{If } h^2 = \tfrac{1}{4} \text{ and } c_{11'} = 1/16 \text{ ,}$$

$$b = \frac{2p}{p + 15 + \frac{\Sigma n_i (n_i-1)}{p}} \quad \text{as compared to } b = \frac{2p}{p+15} \quad \text{assuming no}$$

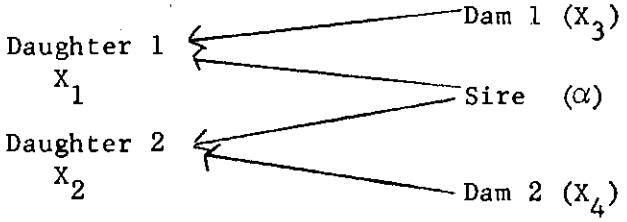
environmental correlation. As before,

$$r_{TI} = \sqrt{\tfrac{1}{2} b} = \sqrt{\frac{p}{p+15 + \frac{\Sigma n_i (n_i-1)}{p}}} \quad \text{for } h^2 = \tfrac{1}{4} \text{ and } c_{11'} = 1/16 \text{ .}$$

Correction for level of mates

If the 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 would be to set up one equation for each daughter record and one equation for each dam record. For two dams and two daughters we would have:



The equations to find the b's for $I = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4$ are:

$$b_1 + \tfrac{1}{4}h^2 b_2 + \tfrac{1}{2}h^2 b_3 + 0\, b_4 = \tfrac{1}{2} h^2$$

$$\tfrac{1}{4}h^2 b_1 + b_2 + 0\, b_3 + \tfrac{1}{2}h^2 b_4 = \tfrac{1}{2} h^2$$

$$\tfrac{1}{2}h^2 b_1 + 0\, b_2 + b_3 + 0\, b_4 = 0$$

$$0 b_1 + \tfrac{1}{2}h^2 b_2 + 0\, b_3 + b_4 = 0$$

As expected $b_1 = b_2 = b$. But, $b_3 = b_4 = -\frac{1}{2} h^2 b$. Thus, the weight for the dam's record is $-\frac{1}{2} h^2$ of that for the progeny.

This is certainly different from the usual daughter-dam comparison where:

Sire value = daughter average - dams' average .

With such a procedure, $b_2 = -b_1$ rather than $-\frac{1}{2} h^2 b_1$.

The equal parent or American index also weights the dam record too much. The "logic" for the method is that

Progeny value = $\frac{1}{2}$ Sire value + $\frac{1}{2}$ Dam value.

Rearrangement of the terms gives

Sire value = 2 times Progeny average value - Dams' average.

so that $b_2 = -\frac{1}{2} b_1$ rather than $b_2 = -\frac{1}{2} h^2 b_1$.

The correct procedure can be simplified so that only 2 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:

$$\frac{1 + (p-1)\, a_{11'} h^2}{p} b_1 + \frac{a_{12} h^2}{p} b_2 = a_{1\alpha} h^2$$

$$\frac{a_{12} h^2}{p} b_1 + \frac{1}{p} b_2 = a_{2\alpha} h^2 .$$

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 covariance zero with the other p-1 dams.

Usually $a_{11'} = \frac{1}{4}$, $a_{12} = \frac{1}{2}$, $a_{1\alpha} = \frac{1}{2}$ and $a_{2\alpha} = 0$.

Then, $b_1 = \frac{\frac{1}{2}ph^2}{1 + (p-1)\frac{1}{4}h^2 - (\frac{1}{2}h^2)^2} = \frac{2p}{p + \frac{4-h^2}{h^2} - h^2}$ and $b_2 = -\frac{1}{2}h^2 b_1$.

Note the similarity of b_1 to the b when dams are not considered. For example, if $h^2 = \frac{1}{4}$

$$b_1 = \frac{2p}{p + 14.75} \quad \text{rather than} \quad b = \frac{2p}{p + 15} .$$

Similarly, the r_{TI} changes only slightly since $a_{2\alpha} = 0$. If $h^2 = 1/4$,

$$r_{TI} = \sqrt{\frac{2p}{p + 14.75}(\tfrac{1}{2}) + 0} = \sqrt{\frac{p}{p + 14.75}} \quad \text{rather than}$$

$$r_{TI} = \sqrt{\frac{p}{p + 15}} \quad \text{when the dams are not considered.}$$

Progeny with different numbers of records

Often in evaluation of sires the progeny will have different numbers of records. A common example is that dairy cows may often have more than one 305-day lactation record.

One solution to the problem of weighting these records would be to set up one equation for each record. Then the weights would be found for each record.

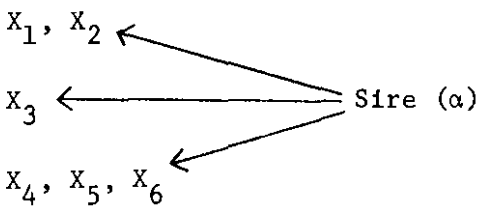
If the simplified equations are used, the diagonal coefficients will be 1. The RHS's will be $a_{i\alpha}h^2$ as before for all i. In the case of half sibs for sire evaluation, these will all be $\frac{1}{2}h^2$. The off-diagonal coefficients will be of two kinds. The coefficients corresponding to covariances among records on the same animal will be repeatability, r, since the covariance, $\sigma_{XX'} = r\sigma_X^2$. The other coefficients will be $a_{ij}h^2$ as before where a_{ij} is the relationship between the animals that made the records. In sire evaluation from half-sister records, these will all be $\frac{1}{4}h^2$.

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 .

Estimate the additive genetic value of their sire from



$$I = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 .$$

The equations to find the b's are:

$$b_1 + rb_2 + \tfrac{1}{4}h^2b_3 + \tfrac{1}{4}h^2b_4 + \tfrac{1}{4}h^2b_5 + \tfrac{1}{4}h^2b_6 = \tfrac{1}{2}h^2$$

$$rb_1 + b_2 + \tfrac{1}{4}h^2b_3 + \tfrac{1}{4}h^2b_4 + \tfrac{1}{4}h^2b_5 + \tfrac{1}{4}h^2b_6 = \tfrac{1}{2}h^2$$

$$\tfrac{1}{4}h^2b_1 + \tfrac{1}{4}h^2b_2 + b_3 + \tfrac{1}{4}h^2b_4 + \tfrac{1}{4}h^2b_5 + \tfrac{1}{4}h^2b_6 = \tfrac{1}{2}h^2$$

$$\tfrac{1}{4}h^2b_1 + \tfrac{1}{4}h^2b_2 + \tfrac{1}{4}h^2b_3 + b_4 + rb_5 + rb_6 = \tfrac{1}{2}h^2$$

$$\tfrac{1}{4}h^2b_1 + \tfrac{1}{4}h^2b_2 + \tfrac{1}{4}h^2b_3 + rb_4 + b_5 + rb_6 = \tfrac{1}{2}h^2$$

$$\tfrac{1}{4}h^2b_1 + \tfrac{1}{4}h^2b_2 + \tfrac{1}{4}h^2b_3 + rb_4 + rb_5 + b_6 = \tfrac{1}{2}h^2$$

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

$I = \Sigma b_iX_i$ are:

$$d_1b_1 + a_{12}h^2b_2 + \ldots + a_{1N}h^2b_N = a_{1\alpha}h^2$$

$$a_{12}h^2b_1 + d_2b_2 + \ldots + a_{2N}h^2b_N = a_{2\alpha}h^2$$

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

$$a_{1N}h^2b_1 + a_{2N}h^2b_N + \ldots + d_Nb_N = a_{N\alpha}h^2.$$

If all animals are half-sibs,

$$d_i = \frac{\frac{1 + (n_i - 1)\, r}{n_i} + (p_i - 1)\tfrac{1}{4}h^2}{p_i} \quad ,$$

$$a_{ij}h^2 = \tfrac{1}{4}\, h^2 \quad , \text{ and}$$

$$a_{i\alpha}h^2 = \tfrac{1}{2}\, h^2 \quad .$$

The r_{TI} will be computed as usual.

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. Ordinarily each female will produce only one set of progeny. The animals in each group will be related as full sibs ($a_{ii'} = \frac{1}{2}$) but will be related as paternal half sibs ($a_{ij} = \frac{1}{4}$) to animals in other groups.

If p_i is the number in each full sib group and $n_i = 1$ and the sire is to be evaluated, the equations defining the b's are:

$$\begin{aligned}
d_1 b_1 + \tfrac{1}{4}h^2 b_2 + \ldots + \tfrac{1}{4}h^2 b_N &= \tfrac{1}{2}\, h^2 \\
\tfrac{1}{4}h^2 b_1 + d_2 b_2 + \ldots + \tfrac{1}{4}h^2 b_N &= \tfrac{1}{2}\, h^2 \\
\vdots \qquad\qquad\qquad\qquad & \quad \vdots \\
\tfrac{1}{4}h^2 b_1 + \tfrac{1}{4}h^2 b_2 + \ldots + d_N b_N &= \tfrac{1}{2}\, h^2 \quad \text{where}
\end{aligned}$$

$d_i = \dfrac{1 + (p_i - 1)\, \frac{1}{2}\, h^2}{p_i}$. Modifications would, of course, have to be made for some $n_i > 1$, and for other possible relationships and environmental correlation which is very likely for animals in the same litter.

Summary IX

PROBABILITY STATEMENTS ABOUT TRUE VALUES

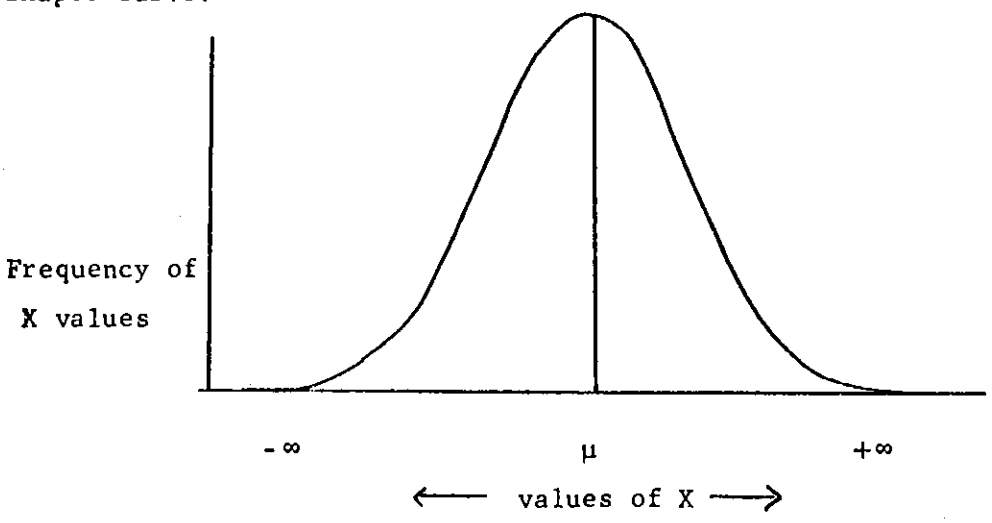
We know that the average true additive genetic value, T, for animals with the same index value, $I = I_o$, is I_o. Thus, I_o is the mean of a distribution of T values for animals with the same index, I_o. The variance of T for $I = I_o$ depends on r_{TI} and σ_T^2 .

$$\sigma^2_{T|I=I_o} = (1-r_{TI}^2)\ \sigma_T^2 \quad .$$

If T and I follow a bivariate normal distribution, I_o and $\sigma^2_{T|I=I_o}$ determine the distribution of T for $I = I_o$. We will see after a discussion of the normal distribution how we can use these to make probability statements about T for $I = I_o$.

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.



We will let X be the set of values having the normal distribution. The mean is also the median of the X values, i.e., half the values are greater than μ and 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 tells us the fraction of the area between μ and $\mu + t\sigma$ or equivalently between μ and $\mu - t\sigma$ since the distribution curve is symmetrical. This fraction corresponds to the probability that a 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 the table (following page):

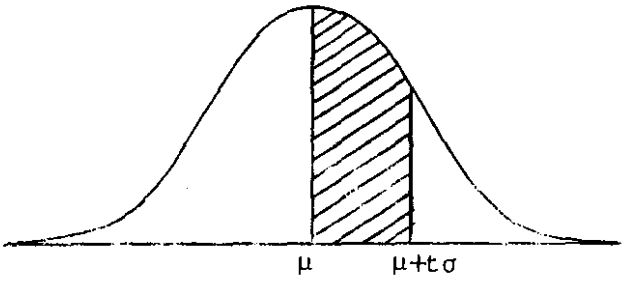
A.) To find probabilities (fractions of total area) corresponding to truncation points which can be expressed as $\mu + t\sigma$ or $\mu - t\sigma$ depending on which side of μ the truncation point is located.

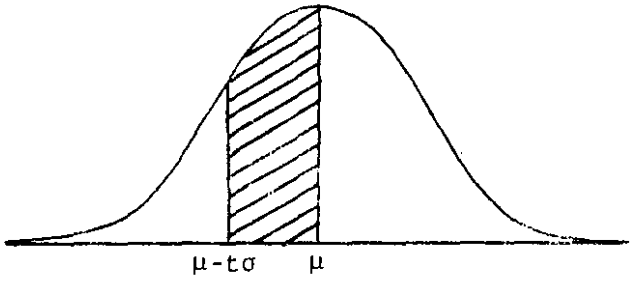
B.) To find truncation points expressed as $\mu + t\sigma$ or $\mu - t\sigma$ corresponding to required probabilities.

Examples of A.

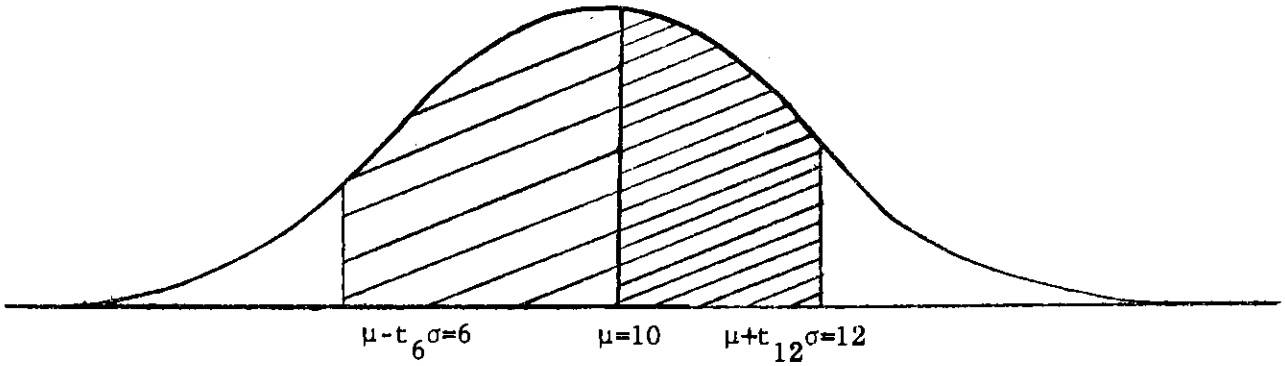
Let $\sigma = 2$ and $\mu = 10$ for distribution of X. Find the probability that a value of X will fall between 6 and 12.

Table of areas under the normal curve.





Truncation Point t	Area between μ+tσ and μ or between μ-tσ 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



Since the table gives the area between μ and $\mu+t\sigma$, we have to find the area between $\mu = 10$ and $\mu + t_{12}\sigma = 12$ and the area between $\mu = 10$ and $\mu - t_6\sigma = 6$. (The subscripts on the t's identify the truncation points.) The total area will be the sum of the two parts. In more formal terms:

$$P(12 > X > 6) = P(10 > X > 6) + P(12 > X > 10) .$$

To use the tables we must find t_{12} and t_6. We know

$\mu + t_{12}\sigma = 12$ but $\mu = 10$ and $\sigma = 2$. Thus

$10 + t_{12}(2) = 12$ and $t_{12} = (12 - 10) / 2 = 1$.

The corresponding area is .34 .

The general method of finding a t corresponding to a positive truncation point greater than the mean is

$$t = \frac{t\,\sigma + \mu - \mu}{\sigma} = \frac{\text{truncation point} - \mu}{\sigma} .$$

We know $\mu - t_6\sigma = 6$. Thus, $10 - t_6 (2) = 6$ and $-t_6 = (6 - 10)/2 = -2$. $t = 2$ and the corresponding area is .477 between 6 and 10.

For t corresponding to truncation points less than the mean, $t = -\frac{\text{truncation point} - \mu}{\sigma}$. In general, $t = \frac{|\text{truncation point} - \mu|}{\sigma}$.

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 have a value between 6 and 12.

Another problem might be to find the probability of an X value above a truncation point.

Example: $\mu + t\sigma = 12$, $\mu = 10$ and $\sigma = 2$.

We know $P(\infty > X > 10) = .5$. Also, we have found $P(12 > X > 10) = .34$. Thus, $P(\infty > X > 12) = .5 - .34 = .16$.

The probability of X less than 12 can be found by similar logic, i.e.,

$$P(12 > X > -\infty) = P(10 > X > -\infty) + P(12 > X > 10) \quad .$$

Thus, $P(12 > X > -\infty) = .5 + .34 = .84$.

Examples of B

1. 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 are usually chosen so that they are symmetrical about μ. In that case:

$\mu + t\sigma$ is the upper limit, and

$\mu - t\sigma$ is the lower limit .

t will be the same in both.

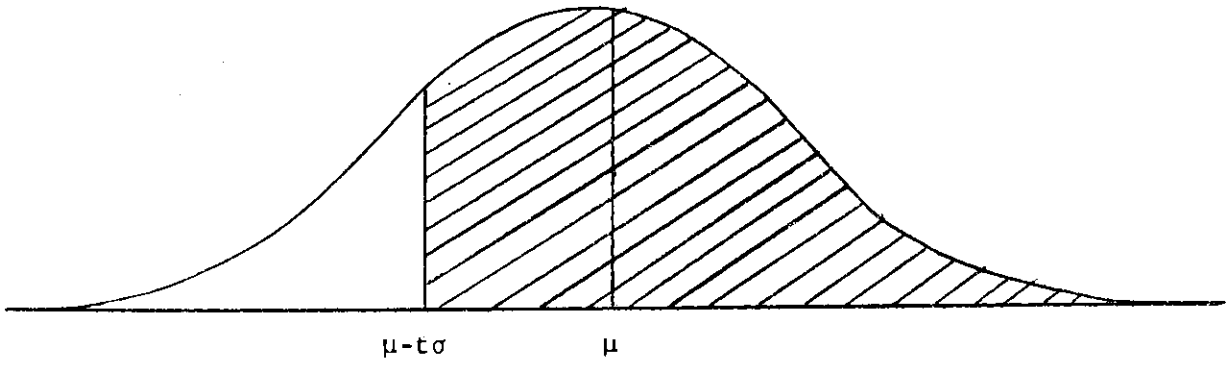
We know the area from μ to $\mu + t\sigma$ must be .90/2 = .45. The t corresponding to an area of .45 is about halfway between 1.6 and 1.7 so let t = 1.65.

If $\mu = 10$ and $\sigma = 2$,

the upper limit is $10 + 1.65(2) = 13.30$, and

the lower limit is $10 - 1.65(2) = 6.70$.

2. Find the truncation point which 90% of the values of X will exceed. $\mu = 10$, $\sigma = 2$.

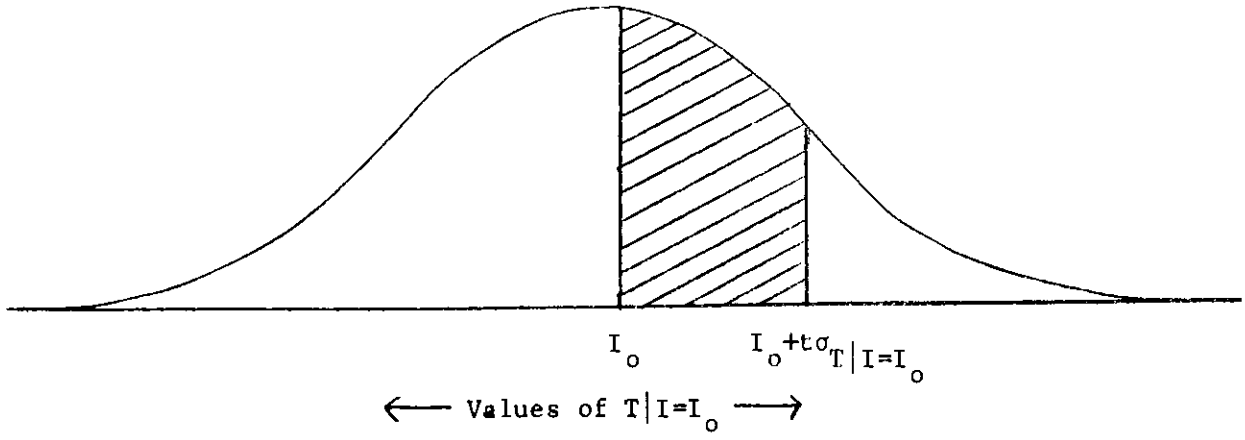


We must find the t corresponding to an area of .40 between μ and μ - tσ. From the table t ≅ 1.3 .

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

We have seen that $\mu_{T|I=I_o} = I_o$. Thus, I_o corresponds to the mean of the distribution of T values for animals with the same index. We will substitute I_o for μ of the general discussion. Similarly we will substitute $\sigma_{T|I=I_o}$ for σ of the general discussion.



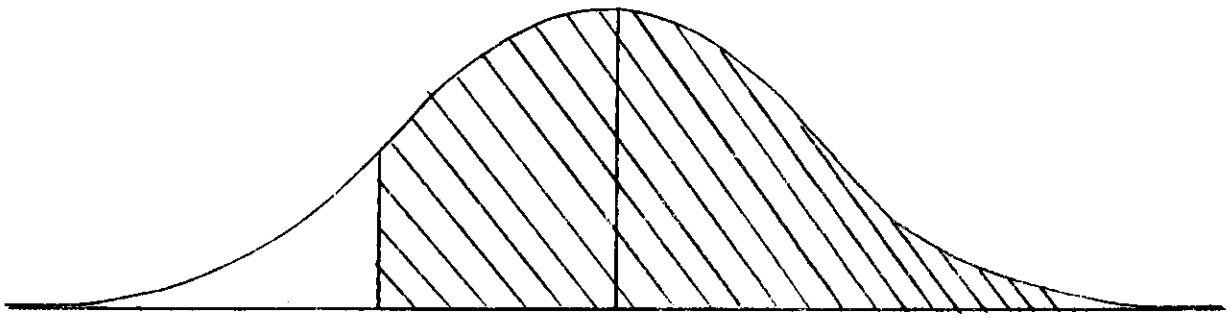
1. Probability statements about genetic values: $T = G_{A_\alpha}$, I predicts additive genetic value.

We have a bull that has 35 progeny with 1 record each averaging +200; $h^2 = 1/4$; $\sigma^2_{10} = 1000000$. What is the probability his true additive genetic value is greater than 0?

$$b = \frac{2p}{p + 15} = \frac{70}{50} \qquad I_o = \frac{70}{50}(200) = 280 .$$

$$r^2_{TI} = \frac{p}{p + 15} = \frac{35}{50} . \qquad \sigma^2_{T|I=I_o} = \left(1 - \frac{35}{50}\right) 1000000 = 300000 .$$

$$\sigma_{T|I=I_o} = 548 .$$



Then, $t = \frac{|0 - 280|}{548} \sim .5$. The corresponding area gives the fraction between 0 and 280 as .19. Thus, the probability of T for the bull exceeding 0 is .5 + .19 = .69. Correspondingly there is a probability 1 - .69 = .31 , that is his true value is less than 0.

The 95% confidence limits on the true value for this bull would be 280 - t(548) to 280 + t(548) . t ~2.0 for area .95 / 2 = .475 .

the upper limit is 280 + 1096 = 1376 , and

the lower limit is 280 - 1096 = -816 .

2. Probability statements about a record.

We previously discussed the probability that an animal's genetic value was between, above, below, etc., certain truncation points given the index estimate and the corresponding r_{TI} and σ^2_G, but in fact when the animal actually makes a record, in addition to its genetic value, a new random environmental effect influences the record. Thus, the variance of records for animals with a predicted genetic value depends on the variance of genetic values given the index plus the variance of environmental effects.

In this case, $T = X_\alpha = G_{A_\alpha} + E_\alpha$, where I predicts 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

$$\sigma_{X_i T} = E(X_i, G_{A_\alpha} + E_\alpha) = E[(G_{A_i} + E_i)(G_{A_\alpha} + E_\alpha)]$$
$$= E(G_{A_i} G_{A_\alpha} + G_{A_i} E_\alpha + G_{A_\alpha} E_i + E_i E_\alpha) .$$

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 may not be zero. If there is no environmental covariance, the right-hand sides are $a_{i\alpha}\sigma^2_{10} = a_{i\alpha}\hat{h}\sigma^2_X$ as for predicting additive genetic value and the index for predicting a future record is exactly the same as for predicting additive genetic value. The reason, of course, is that there is no way of predicting E_α for the new record.

The r_{TI} and σ^2_T, however, are different from when I predicted G_{A_α}. $\sigma^2_T = E(T^2) = E(G_{A_\alpha} + E_\alpha)^2 = \sigma^2_{G_A} + \sigma^2_E = \sigma^2_X$ rather than $\sigma^2_{G_A} = \hat{h}\sigma^2_X$. The numerator of r_{TI} is $\Sigma b_i a_{i\alpha}\hat{h}\sigma^2_X$ as before, but

$$r_{TI} = \sqrt{\frac{\Sigma b_i a_{i\alpha} h^2 \sigma_X^2}{\sigma_X^2}} = \sqrt{h^2 \Sigma b_i a_{i\alpha}}$$

rather than $\sqrt{\Sigma b_i a_{i\alpha}}$ since prediction of E_α is zero. Then $\sigma^2_{T|I=I_o} = (1-h^2\Sigma b_i a_{i\alpha})\sigma_X^2$ rather than $(1-\Sigma b_i a_{i\alpha})h^2\sigma_X^2$. Notice that many of the same quantities, $\Sigma b_i a_i$, h^2, σ_X^2, are involved whether prediction is for G_{A_α} or $X = G_{A_\alpha} + E_\alpha$; the arrangement, however, is different.

Example. 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, the sire's and dam's estimated genetic value are used in estimating the genetic value of their progeny: $\hat{G}_{progeny} = \frac{\hat{G}_{sire} + \hat{G}_{dam}}{2}$.

The r^2_{TI} for progeny = $1/4r^2_{TI}$ for sire + $1/4r^2_{TI}$ for dam if sire and dam are unrelated. Assume for milk yield $h^2 = 1/4$ and $\sigma_X^2 = (2000\ \text{lb})^2$. The following table shows the effect of increasing r^2_{TI} for sire and dam on $\sigma_{X|I=I_o}$.

r^2_{TI}			$\sigma_{X\|I=I_o}$	95% C.I. =
Sire	Dam	Progeny	(for progeny)	$I_o \pm 1.96\ \sigma_{X\|I=I_o}$
0	0	0	2000 lb.	$I_o \pm$ 3920
.25	0	.0625	1984	$I_o \pm$ 3889
.25	.25	.1250	1968	$I_o \pm$ 3857
.50	.50	.2500	1936	$I_o \pm$ 3795
.75	.50	.3125	1920	$I_o \pm$ 3763
.75	.75	.3750	1904	$I_o \pm$ 3732
1.00	.75	.4375	1887	$I_o \pm$ 3699
1.00	1.00	.5000	1871	$I_o \pm$ 3667

The obvious conclusion from this chart is that the average error of predicting a record ($\sigma_{X|I=I_o}$) does not decrease very much even with perfect prediction of the parents' genetic values when σ_E^2 is relatively large.

3. Probability statements about differences in genetic values for animals with indexes I_1 and I_2.

Suppose 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 genetic values for animals with index values I_1 and I_2 will have a distribution. The immediate problem is to determine the mean and variance for the distribution of $T_1-T_2|I_1-I_2$. The mean is the same as the mean of $(T_1|I = I_1)-(T_2|I = I_2)$ 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.

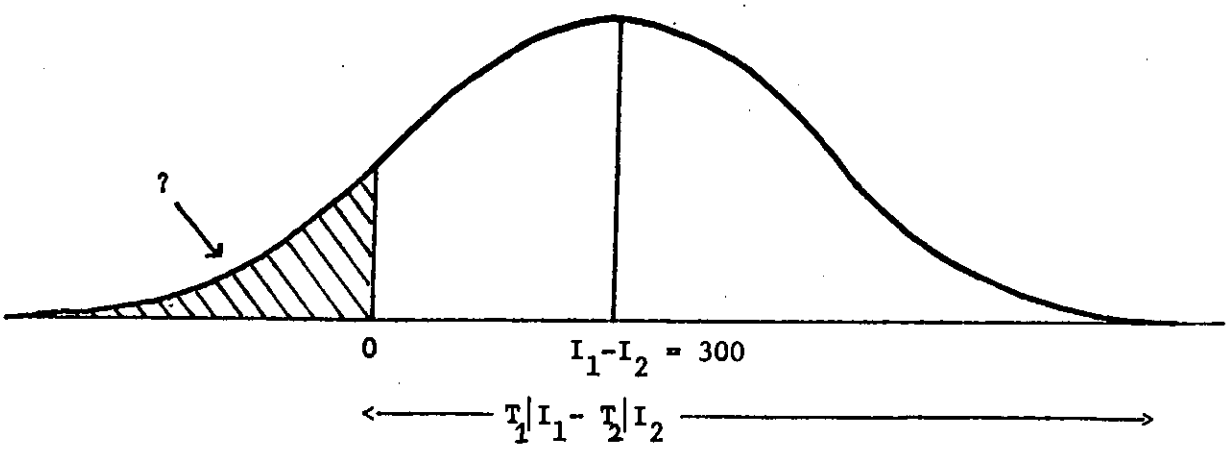
$$V[(T_1|I=I_1) - (T_2|I=I_2)] = V(T_1|I=I_1) + V(T_2|I=I_2)$$

since 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.

Example: Suppose $I_1 = 500$ and $I_2 = 200$, i.e., $I_1-I_2 = 300$ and $r^2_{TI_1} = 3/4$ and $r^2_{TI_2} = 1/4$ and $\sigma^2_G = (1000)^2$. What is the probability that the true difference in genetic values is 0 or less (i.e., animal with $I_2 = 200$ actually has equal or greater true value than animal with $I_1 = 500$?



We want to find $\text{Prob}[(T_1|I_1 - T_2|I_2) < 0]$.

Thus $0 = (I_1 - I_2) - t\sqrt{2 - r^2_{TI_1} - r^2_{TI_2}}\,\sigma_G$

and $t = (300 - 0)/[\sqrt{2 - 3/4 - 1/4}\ \ (1000)] = .3$.

The corresponding area between 0 and 300 is .12 and the area below 0 is .5 - .12 = .38, which is the probability of the animal with the lower index, $I_2 = 200$, actually having a higher true value than the animal with the higher index, $I_1 = 500$.

A more direct approach would be to define $T = T_1 - T_2$ and use all information to predict T and then follow the general selection index procedure.

Summary of various distributions associated with the 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 the development of the selection index for a particular trait there are at least 6 variables.

1) The basic distribution is of the phenotypic records. The P's, or as we have also called them, 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) If we are attempting to evaluate additive genetic values, the mean is $\mu_G = 0$ and the variance is $\sigma^2_G = \sigma^2_{10} = h^2\sigma^2_X$. Note that $\sigma^2_G \leq \sigma^2_X$ since $h^2 \leq 1$.

3) Our criterion for evaluating the G's is the index estimate, I. The mean is $\mu_I = 0$ and the variance is $\sigma_I^2 = r_{TI}^2\sigma_G^2$. Note that $\sigma_I^2 \leq \sigma_G^2$ since $r_{TI}^2 \leq 1$.

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 $\sigma^2_{T|I=I_o} = (1-r_{TI}^2)\sigma_T^2$; [$= (1-r_{TI}^2)\sigma_G^2$ 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 $\sigma^2_{X|I=I_o} = (1-r_{TI}^2)\sigma_G^2 + \sigma_E^2$ when r_{TI}^2 is for predicting G or $= (1-r_{TI}^2)\sigma_X^2$ when r_{TI}^2 is for predicting X.

6) The difference in 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 $\sigma^2_{T|I=I_1} + \sigma^2_{T|I=I_2} = (2-r^2_{TI_1}-r^2_{TI_2})\sigma_G^2$.

Corresponding to the general σ, the standard deviations for the six distributions are:

1) X's , $\sigma_X = \sqrt{\sigma_G^2 + \sigma_E^2}$;

2) G's , $\sigma_G = \sqrt{h^2}\sqrt{\sigma_G^2 + \sigma_E^2}$;

3) I's , $\sigma_I = r_{TI}\sigma_G$;

4) G's given $I=I_o$, $\sigma_{T|I=I_o} = \sqrt{1-r_{TI}^2}\ \sigma_G$;

5) X's given $I=I_o$, $\sigma_{X|I=I_o} = \sqrt{(1-r_{TI}^2)\sigma_G^2 + \sigma_E^2} = \sqrt{\sigma_G^2 + \sigma_E^2} \cdot \sqrt{1-r_{TI}^2 h^2}$ when r_{TI}^2 is for predicting G ;

6) $G_1|I=I_1 - G_2|I=I_2$, $\sigma_{G_1-G_2|I_1-I_2} = \sqrt{2-r^2_{TI_1}-r^2_{TI_2}}\ \sigma_G$.

Summary X

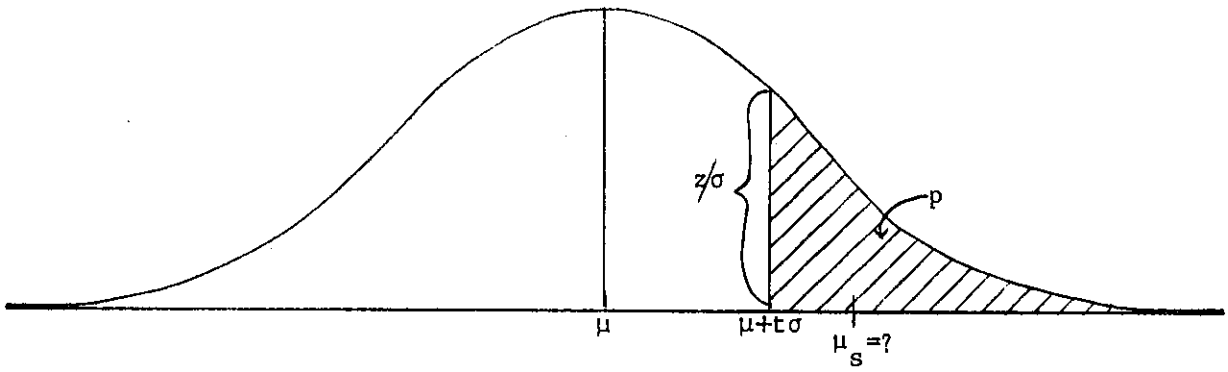
SUPERIORITY IN T OF SELECTED GROUP

Average of selected group

The basic principle in selection is to select the best and cull the rest. We have said the selection index is the best method of evaluating animals to determine which to select or cull. How can we determine how much better the selected ones are expected to be than the original group?

The normal distribution

The basic problem is this. If a fraction, p, are 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 diagramed as:



The truncation point, $\mu+t\sigma$, depends on p as before.

The expected or average value of the fraction p can be found from:

$$\mu_s = \frac{1}{p} \int_{\mu+t\sigma}^{\infty} xf(x)dx = \mu+D\sigma$$ where f(x) is the density function of the normal distribution.

$D\sigma$ 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.

Note that $\mu_s - \mu = D\sigma$, which is sometimes known as the selection differential. If $\mu = 0$, $\mu_s = D\sigma$.

The table of D for small samples is based on order statistics. The values are not the same as z/p. The table of D for large samples is the same as z/p. Dr. C. R. Henderson has 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

A breed organization reports a 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. What would we expect the average of the 100 to be?

Fraction selected = 20/100 = .20. The corresponding D = 1.4.

$D' = 1.4 - .25/20 = 1.3875$.

We know

$$\mu_s = \mu + D\sigma ,$$

$$1000 = \mu + 1.3875(2000), \quad \mu = 1000 - 1.3875(2000) = -1775 \text{ lb} .$$

We would have been misled considerably if we had evaluated the bull on his top 20 daughters.

Question. What should be the number of daughters to use in the formula for estimating the genetic value of this bull--20 or 100?

Table of D (small samples)

Expected average of a group selected out of a sample from a normal population when the sample size is small (in units of $\sigma = 1$)

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 of D (large samples)

Expected Average of Certain Fractions Selected Out of a Sample from a Normal Population (in units of $\sigma = 1$)

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.

Schaeffer (Biometrics, Dec. 1970) has developed a solution for this problem which generally depends only on the fraction selected.

Genetic superiority of selected group

We have estimated T from I and have selected a fraction of animals based on their index values. What will be the expected superiority in T of the selected group?

The selected I's 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, we have $\mu_{T_s} = \mu_T + r_{TI}\sigma_T D$ as we stated before. The same result can be obtained by the regression of T on I:

$$\mu_{T_s} = \mu_T + b_{T\cdot I}(\mu_{I_s} - \mu_I) = \mu_T + (\sigma_{TI}/\sigma_I^2)(\mu_I + D\sigma_I - \mu_I)$$
$$= \mu_T + (\sigma_{TI}/\sigma_I)D = \mu_T + r_{TI}\sigma_T D \text{ by multiplying by } \sigma_T/\sigma_T.$$

Thus, the genetic selection differential will be $\Delta G = r_{TI}D\sigma_T$ per generation. If L is the generation interval in years, then the 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 and in fact the difference in the averages is the selection index prediction of the difference between the selected group and the group they were selected from.

The expression $\Delta G/yr = r_{TI}D\sigma_T/L$ can be used to compare various selection programs. This is the key equation for genetic improvement. Sometimes the best balance of r_{TI}, D, and L will have to be found.

Example:

There are only 1000 progeny available each year for progeny testing. Two replacements are needed each year from the males which are progeny tested. $h^2 = .25$, $\sigma_T = 1000$ lb. milk. The following table illustrates that neither the highest r_{TI} nor the highest selection intensity gives the highest genetic progress.

Some possible combinations of no. males and no. progeny Selected/sampled	%	No. progeny per male sampled	$r_{TI} = \sqrt{\frac{p}{p+15}}$	$D' = D - \frac{.25}{2}$	σ_T	ΔG
2 of 2	100	500	.995	0	1000 lb	0 lb
2 of 5	40	200	.964	.84	1000	815
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

These ΔG values suggest that of the six combinations, testing 50 males with 20 progeny each is best. In actual practice, income and cost values must be assigned to each plan. Since ΔG for 2 of 20 is nearly as great as ΔG for 2 of 50, this 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 may all 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, since progeny receive a sample half of the genetic value of each of their parents, $G_{progeny} = (\Delta S + \Delta D)/2$.

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 $[(\Delta S/L_S) + (\Delta D/L_D)]/2$. Proof:

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 bulls born L_S years ago is $P - L_S\Delta g$. The superiority of the selected bulls over that average is ΔS. Thus, $S = P - L_S\Delta g + \Delta S$. Similarly, $D = P - L_D\Delta g + \Delta D$. We know $P = (S + D)/2$ so that by substitution $P = (S + D)/2 = (1/2)(P - L_S\Delta g + \Delta S + P - L_D\Delta g + \Delta D)$. Then,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 $\Delta g = (\Delta S + \Delta D)/(L_S + L_D)$.

This result is due to Dickerson and Hazel (1944).

Rendel and Robertson (1950) have extended this procedure to consider genetic value of sires of sires (SS), dams of sires (DS), sires of dams (SD), and dams of dams (DD) selected as grandparents each with different generation intervals (L_{SS}, L_{DS}, L_{SD}, and L_{DD}, respectively).

$$E(X_iT) = \sigma_{X_iT} = Cov(X_i, v_1G_1 + \ldots + v_MG_M)$$
$$= v_1\sigma_{G_iG_1} + v_2\sigma_{G_iG_2} + \ldots + v_M\sigma_{G_iG_M}$$
$$= \sum_{j=1}^{M} v_j\sigma_{G_iG_j}$$

Recall that $\sigma_{G_iG_j} = r_{g_{ij}}\sqrt{h_i^2h_j^2}\,\sigma_{X_i}\sigma_{X_j}$ and also note that when $G_i = G_j$, then $\sigma_{G_iG_i} = \sigma^2_{G_i} = h_i^2\sigma^2_{X_i}$. Solving the equations for the β's then gives $I = \beta_1X_1 + \beta_2X_2 + \ldots + \beta_NX_N$ which is the same index as found when indexing each trait separately and then weighting by economic value as $I = v_1I_1 + \ldots + v_MI_M$.

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 $\Delta T = r_{TI}D\sigma_T$ where σ^2_I can be found as the variance of the linear function

$$V(I) = E[(\sum_{i=1}^{N} \beta_iX_i)^2] \text{ and}$$

$r_{TI} = \sqrt{\Sigma\beta_i\sigma_{X_iT}/\sigma^2_T}$ where σ_{X_iT} (i=1, ..., N) are the covariances of linear functions and σ^2_T is the variance of the linear function, $T = \sum_{j=1}^{M} v_jG_j$.

Often the expected correlated response for one or more traits is of interest when selecting for some overall economic value. For any index, whether the selection index or any other, the correlated genetic response for any trait j can be found by the regression of G_j on I: $\hat{G}_j = \mu_{G_j} + b_{G_j.I}(I_{sel} - \mu_I)$, where $I_{sel} = \mu_I + D\sigma_I$ and $\hat{G}_j = \mu_{G_j} + \Delta G_j$; thus, $\Delta G_j = [Cov(G_j,I)]D/\sigma_I$ where $Cov(G_j,I) = Cov(G_j, \beta_1X_1 + \ldots + \beta_NX_N) = \beta_1\sigma_{G_jG_1} + \beta_2\sigma_{G_jG_2} + \ldots + \beta_N\sigma_{G_jG_N}$. This formula holds for any trait whether included in T or I. However, the correlated responses of the traits included in T weighted by their economic values will equal total economic response; i.e., $\Delta T = v_1\Delta G_1 + v_2\Delta G_2 + \ldots + v_M\Delta G_M$.

An example follows for selection for two traits. Included are examples of comparing correlated response in the two traits when selection is for only one of them using either both traits or only one of them.

Example of selecting for more than one trait

Let milk yield = Trait 1, type score = Trait 2

$\sigma^2_{X_1} = (2000\ \underline{lb})^2$ $\sigma^2_{X_2} = (2\%)^2$ $\sigma_{X_1X_2} = 400\ \underline{lb}\ \underline{\%}$ $r_p = .1$

$\sigma^2_{G_1} = (1000\ \underline{lb})^2$ $\sigma^2_{G_2} = (1\%)^2$ $\sigma_{G_1G_2} = 200\ \underline{lb}\ \underline{\%}$ $r_g = .2$

$h^2_1 = 1/4$ $h^2_2 = 1/4$

Suppose $v_1 = \$.025/\underline{lb}$ $v_2 = \$50./\%$

Method 1. Find $I_1 = b_{11} X_1 + b_{12} X_{12}$

$$4{,}000{,}000\ b_{11} + 400\ b_{12} = \sigma_{X_1G_1} = \sigma^2_{G_1} = 1{,}000{,}000$$

$$400\ b_{11} + 4\ b_{12} = \sigma_{X_2G_1} = \sigma_{G_2G_1} = 200$$

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 $I_2 = .00002525\ X_1 + .2475\ X_2$.

$I = .025\ I_1 + 50.\ I_2 = [.025(.2475) + 50(.00002525)]X_1 + [.025(25.2525) + 50(.2475)]X_2$

$I = \beta_1 X_1 + \beta_2 X_2 = .00745\ X_1 + 13.006\ X_2$ will be the overall index.

Total response: $\triangle G = D\ \sigma_I$; $\sigma^2_I = \beta^2_1 \sigma^2_{X_1} + \beta^2_2 \sigma^2_{X_2} + 2\ \beta_1\beta_2\ \sigma_{X_1X_2} = 976$

$\triangle G = D\ \sqrt{976} = 31.24\ D$ ($), total expected response.

Correlated responses:

$$\triangle_c G_2 = \frac{Cov(G_2, I)}{\sigma_I} D\ ;\quad Cov(G_2, I) = \beta_1\ \sigma_{G_2G_1} + \beta_2\ \sigma^2_{G_2} = .00745(200)$$

$+ 13.066(1) = 14.5$; thus $\triangle_c G_2 = \frac{14.5}{31.24} D = .464\%\ (D)$.

$$\triangle_c G_1 = \frac{Cov(G_1, I)}{\sigma_I} D\ ;\quad Cov(G_1, I) = \beta_1\ \sigma^2_{G_1} + \beta_2\ \sigma_{G_1G_2} = 10051;$$

Thus $\triangle_c G_1 = \frac{10051}{31.24} D = 321.7\ \underline{lb}\ (D)$.

$\triangle G$ should $= v_1\ \triangle_c G_1 + v_2\ \triangle_c G_2 = .025(321.7\ D) + 50.\ (.464\ D) =$

$8.04\ D + 23.2\ D = 31.24\ D$.

Suppose another trait, e.g., fat test = Trait 3, is of interest, then

$$\Delta_c G_3 = \frac{\mathrm{Cov}(G_3,I)}{\sigma_I} D; \quad \mathrm{Cov}(G_3,I) = \beta_1 \sigma_{G_3G_1} + \beta_2 \sigma_{G_3G_2}.$$

If $r_{g_{13}} = -.6$, $r_{g_{23}} = .1$, $\sigma^2_{X_3} = (.3\%^F)^2$, $h^2_3 = .5$; then $\sigma^2_{G_3} = .045\ \%^2$

$$\sigma_{G_1G_3} = r_{g_{13}} \sqrt{\sigma^2_{G_1}\ \sigma^2_{G_3}} = -.60\ \sqrt{(1000)^2(.045)} = -127 \text{ lb } \%^F, \text{ and}$$

$$\sigma_{G_2G_3} = r_{g_{23}} \sqrt{\sigma^2_{G_2}\ \sigma^2_{G_3}} = .10\ \sqrt{(1)^2\ (.045)} = .02121\ \%^T\%^F.$$

Then $\mathrm{Cov}(G_3,I) = .00745(-127) + 13.006(.02121) = -.67.$

Thus $\Delta_c G_3 = \frac{-.67}{31.24} D = -.021\ D\ (\%^F).$

Example of selecting for one trait using both traits

Suppose $v_2 = 0$, then v_1 can be any positive nonzero value; 1 is convenient, i.e., $I = v_1 I_1$ and obviously v_1 will not change ranking.

If $v_1 \neq 1$ then $\Delta_c G_1 = \frac{\mathrm{Cov}(G_1,I)}{\sigma_I} D$ but $I = v_1 I_1$ so that

$$\Delta_c G_1 = \frac{v_1\ \mathrm{Cov}(G_1,I_1)}{v_1\ \sigma_{I_1}} D = \frac{\mathrm{Cov}(G_1,I_1)}{\sigma_{I_1}} D \quad \text{as for } v_1 = 1.$$

Thus for $v_1 = 1$, $v_2 = 0$: $I = I_1 = .2475\ X_1 + 25.2525\ X_2$

Response: $\Delta G_1 = \Delta I_1 = D\ \sigma_{I_1}$; $\sigma^2_{I_1} = (.2475)^2\sigma^2_{X_1} + (25.2525)^2\sigma^2_{X_2}$

$$+ 2(.2475)\ (25.2525)\ \sigma_{X_1X_2},$$

$$\sigma^2_{I_1} = 250{,}556 \quad \text{and} \quad \sigma_{I_1} = 500.56\ .$$

Thus $\Delta G_1 = 500.56\ D$ (lb)

Correlated response: When selecting for G_1 using X_1 and X_2

$$\Delta_c G_2 = \frac{\mathrm{Cov}(G_2,I_1)}{\sigma_{I_1}} D = \frac{b_{11}\ \sigma_{G_2G_1} + b_{12}\ \sigma^2_{G_2}}{\sigma_{I_1}} D = \frac{74.75}{500.56} D = .149\ D\ (\%)$$

$$\Delta_c G_3 = \frac{\mathrm{Cov}(G_3,I_1)}{\sigma_{I_1}} D = \frac{b_{11}\ \sigma_{G_3G_1} + b_{12}\ \sigma_{G_3G_2}}{\sigma_{I_1}} D = \ ?$$

Comparison with selecting for one trait using only record of that trait

Suppose $I_1 = b_1 X_1 = h_1^2 X_1 = .25\ X_1$

Response: $\triangle G_1 = D\ \sigma_{I_1}$; $\sigma_{I_1} = \sqrt{(.25)^2 \sigma_{X_1}^2} = 500$; thus $\triangle G_1 = 500\ D$ (1b)

Correlated response: $\triangle G_2 = \frac{Cov(G_2, I_1)}{\sigma_{I_1}} D = \frac{b_1\ \sigma_{G_2 G_1}}{\sigma_{I_1}} D = \frac{.25(200)}{500} D$

$= .1\ D$ (%)

This would be the same response as selecting for trait 2 using only trait 1 since the genetic covariance is positive.

Comparison with selecting for one trait using only record of another trait:

Select for G_2 using X_1 by $I_2 = b_1 X_1$

$\sigma_{X_1}^2\ b_1 = \sigma_{X_1 G_2} = \sigma_{G_1 G_2}$; $b_1 = \frac{\sigma_{G_1 G_2}}{\sigma_{X_1}^2} = \frac{200}{4{,}000{,}000} = .00005$

$I_2 = .00005\ X_1$; $\triangle G_2 = D\ \sigma_{I_2}$; $\sigma_{I_2} = \sqrt{(.00005)^2 \sigma_{X_1}^2} = .1$

Response: Thus $\triangle G_2 = .1\ D$ (%) as above.

Summary

These examples illustrate the method of comparing different selection systems:*

Selection For	Based on	$\triangle G_1$	$\triangle G_2$	$\triangle$ G for .025 = v_1 & 50 = v_2
$.025\ G_1 + 50\ G_2$	X_1, X_2	321.7 1b	.464 %	31.24 $
G_1	X_1, X_2	500.56 1b	.149 %	19.96 $
G_1	X_1	500 1b	.100 %	17.50 $
G_2	X_1	500 1b	.100 %	17.50 $
G_2	X_2	100 1b	.500 %	27.50 $

*All expected responses should be multiplied by D.

An approximate procedure for selecting for more than one trait

Often the genetic correlations needed to find the weights for the index to estimate , $T = v_1G_1 + \ldots + v_MG_M$, are not known or are estimated with not much reliability. In addition the equations to determine the weights are 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 the 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, that is the assumption made to obtain the approximate index.

When only one record is available on each trait of the animal to be evaluated, the indexes for the traits are

$I_j = h_j^2 X_j$ and so the approximate overall index is

$I = v_1h_1^2X_1 + v_2h_2^2X_2 + \ldots + v_Mh_M^2X_M$.

Note that the phenotypic records are weighted by the product of their value and heritability which would be the weights found by solving the equations for the β's when all the phenotypic and genetic correlations are zero:

$$\begin{aligned}
\sigma^2_{X_1}\ \beta_1 + 0\text{'s} &= v_1h_1^2\ \sigma^2_{X_1} + 0\text{'s} \\
\sigma^2_{X_2}\ \beta_2 + 0\text{'s} &= v_2h_2^2\ \sigma^2_{X_2} + 0\text{'s} \\
&\vdots \\
0\text{'s} + \sigma^2_{X_M}\ \beta_M &= 0\text{'s} + v_Mh_M^2\ \sigma^2_{X_M} \ .
\end{aligned}$$

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. If the correct genetic and phenotypic correlations are known the approximate procedure still 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 procedure.

Let $I_A = \beta_1^A X_1 + \ldots + \beta_M^A X_M$ be the approximate index. Then the correlated response for trait j using the approximate index is

$$\Delta G_j = \frac{\mathrm{Cov}(G_j, I_A)}{\sigma_{I_A}} D \quad \text{as before where } E(G_j, I_A) =$$

$$\mathrm{Cov}(G_j, I_A) = \beta_1^A \sigma_{G_j G_1} + \beta_2^A \sigma_{G_j G_2} + \ldots + \beta_M^A \sigma_{G_j G_M} .$$

$\sigma^2_{I_A}$ 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 β'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:

$$4{,}000{,}000\ b_{11}^{A} + 0\ b_{12}^{A} = 1{,}000{,}000$$
$$0\ b_{11}^{A} + 4\ b_{12}^{A} = 0 \quad ; \quad I_{1_A} = .25\ X_1 + 0\ X_2 = h_1^2 X_1 .$$

The equations to find I_{2_A} are:

$$4{,}000{,}000\ b_{21}^{A} + 0\ b_{12}^{A} = 0$$
$$0\ b_{21}^{A} + 4\ b_{22}^{A} = 1 \quad ; \quad I_{2_A} = 0\ X_1 + .25\ X_2 = h_2^2 X_2 .$$

Then $I_A = v_1 I_{1_A} + v_2 I_{2_A} = v_1 h_1^2 X_1 + v_2 h_2^2 X_2 = .025(.25)\ X_1 + 50(.25)\ X_2 .$

$I_A = .00625\ X_1 + 12.5\ X_2$ as compared to the optimum index of

$I = .00745\ X_1 + 13.006\ X_2$.

Total and correlated responses:

Total response computed as usual as

$\triangle T = \sigma_{I_A} D$; $[\sigma_{I_A}^2 = (.00625)^2 \sigma_{X_1}^2 + (12.5)^2 \sigma_{X_2}^2 + 0 = \underline{781.25}, \sigma_{I_A} = 27.95]$ would not be correct since the $\sigma_{X_1X_2}$ really isn't zero. Actually $\sigma_{I_A} = (.00625)^2 \sigma_{X_1}^2 + (12.5)^2 \sigma_{X_2}^2 + 2(.00625)(12.5)\ \sigma_{X_1X_2} = \underline{843.75}$; $\sigma_I = 29.05$. The correct expected total response can be computed as

$\triangle T = \dfrac{\mathrm{Cov}(T, I_A)}{\sigma_{I_A}} D$ where σ_{I_A} is computed using the actual $\sigma_{X_1X_2}$ and $T = v_1 G_1 + v_2 G_2$. The correct expected total response can also be computed from $v_1 \triangle G_1 + v_2 \triangle G_2$.

$$\triangle G_1 = \frac{\mathrm{Cov}(G_1, I_A)}{\sigma_{I_A}} D = \frac{.00625\ \sigma_{G_1}^2 + 12.5\ \sigma_{G_1G_2}}{29.05} D = \frac{8750}{29.05} D = 301.2 \text{ lb (D)}$$

$$\triangle G_2 = \frac{\mathrm{Cov}(G_2, I_A)}{\sigma_{I_A}} D = \frac{.00625\ \sigma_{G_1G_2} + 12.5\ \sigma_{G_2}^2}{29.05} D = \frac{13.75}{29.05} D = .473\ \% \text{ (D)}$$

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: $\Delta G_1 = [\mathrm{Cov}(G_1, I_A)]D/\sigma_{I_A} = (.00625\ \sigma_{G_1})d/27.95 = 223.6$ lb (D) and $\Delta G_2 = [\mathrm{Cov}(G_2, I_A)]D/\sigma_{I_A} = (12.5\ \sigma_{G_2})D/27.95 = .447\%$ (D). Note that the

genetic covariance term in the numerator is ignored and that the incorrect σ_{I_A} is used which ignores the phenotypic covariance.

The three sets of calculated responses which can be compared are summarized below:

1) using the correct covariances,

2) using zero covariances to compute the index but using the correct covariances to compute response, and

3) using zero covariances when really not correct.

Computing of index	response	Expected Response / D $\triangle G_1$ (lb)	$\triangle G_2$ (%)	$\triangle T = v_1 \triangle G_1 + v_2 \triangle G_2$
correct,	correct	321.7	.464	31.24
incorrect,	correct	301.2	.473	31.18
incorrect,	incorrect	223.6	.447	27.95

Using records on all traits of relatives

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 weights.

$P_{11} = G_{11} + E_{11}$ and $P_{23} = G_{23} + E_{23}$ is our usual model. 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 except $\text{Cov}(G_{11}\ G_{23})$ usually will be assumed to be zero. Then the covariance is the covariance between the genetic value for trait 1 on relative 1 and the genetic value for trait 3 on relative 2. If these were measured on the same animal, i.e., relative 1 was relative 2 then the covariance is simply the additive genetic covariance between traits 1 and 2. But in general, the additive genetic covariance is $a_{12}\ \sigma_{G_1G_3}$; the additive relationship between the relatives times the genetic covariance between the traits. This corresponds to the additive genetic covariance between relatives for the same trait, $a_{12}\ \sigma^2_G$.

Thus, if only additive genetic effects are considered $\text{Cov}(G_{ij}\ G_{i'j'}) = a_{ii'}\ \sigma_{G_iG_{j'}}$ 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 also

$\text{Cov}(P_{ij}\ P_{i'j'}) = a_{ii'}\ \sigma_{G_iG_{j'}}$ but when i = i'

$\text{Cov}(P_{ij}\ P_{ij'}) = \sigma_{G_jG_{j'}} + \sigma_{E_jE_{j'}}$ as before and when i = i' 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 to let P_{ij} be 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 $T = \sum_{m=1}^{M} v_m\ G_m$ will be $I = \sum_{k=1}^{N} \beta_k\ X_k$. The equations which determine the β's come as usual from maximizing r_{TI} or minimizing $E(T - I)^2$.

Again, either finding the index directly or weighting the indexes for the economic traits by the economic values are equivalent. We will describe the procedure of finding the index for each trait using all the X's and then putting them together as $I = \sum_{m=1}^{M} v_m\ I_m$.

The basic step then is to estimate $G_{\alpha m}$ the additive genetic value for trait m for animal α from all X's $(X_k,\ k=1,\ \ldots,\ N)$ as

$I_{\alpha m} = b_{m1}\ X_1 + b_{m2}\ X_2 + \ldots + b_{mN}\ X_N$. Note we now must keep track of the relationships among the relative groups and the animal being indexed.

The general equations to find the b's are:

$$\begin{aligned}
&\sigma^2_{X_1} b_{m1} + \sigma_{X_1X_2} b_{m2} + \ldots + \sigma_{X_1X_N} b_{mN} = \sigma_{X_1 G_{\alpha m}} \\
&\vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \vdots \\
&\sigma_{X_1X_N} b_{m1} + \sigma_{X_2X_N} b_{m2} + \ldots + \sigma^2_{X_N} b_{mN} = \sigma_{X_N G_{\alpha m}} .
\end{aligned}$$

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, $\sigma^2_{X_k}$

$$\sigma^2_{X_k} = \sigma^2_{X_{ij}} = \left(\frac{\frac{1+(n_k-1)r_j}{n_k} + (p_k-1)\, a_{ii'} h^2_j}{p_k} \right) \sigma^2_{P_j} \qquad \text{where}$$

$\sigma^2_{P_j}$ 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, $\sigma_{X_kX_{k'}}$

There are three possible types of covariances.

1) If k = ij and k' = i'j (different relative group, same trait j)

$\sigma_{X_kX_{k'}} = \sigma_{X_{ij}X_{i'j}} = a_{ii'} \sigma^2_{G_j}$ 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)

$$\sigma_{X_kX_{k'}} = \sigma_{X_{ij}X_{ij'}} = \frac{\sigma_{P_jP_{j'}} + (p_k-1) a_{ii'} \sigma_{G_jG_{j'}}}{p_k} \qquad \text{where}$$

$\sigma_{P_jP_{j'}}$ 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)

$$\sigma_{X_k X_{k'}} = \sigma_{X_{ij} X_{i'j'}} = a_{ii'} \sigma_{G_j G_{j'}} .$$

Covariances on the RHS's, $\sigma_{X_k G_{\alpha m}}$

The covariances between the X_k and $G_{\alpha m}$ will be of two types:

1) If k = im (same trait as G_m) then

$$\sigma_{X_k G_{\alpha m}} = \sigma_{X_{im} G_{\alpha m}} = a_{i\alpha} \sigma^2_{G_m} \quad \text{and}$$

2) if k = ij (different trait from G_m) then

$$\sigma_{X_k G_{\alpha m}} = \sigma_{X_{ij} G_{\alpha m}} = a_{i\alpha} \sigma_{G_j G_m} .$$

Solving the equations gives $I_{\alpha m} = \sum_{k=1}^{N} b_{mk} X_k$ for trait m. This procedure will be repeated for all economic traits. Note that the coefficients of the b's will be the same for all sets of equations, only the RHS's will change depending on the trait being indexed.

Finally, $I_\alpha = \sum_{m=1}^{M} v_m I_{\alpha m} = \sum_{k=1}^{N} \beta_k X_k$.

Response from selection

As usual $\triangle T = D \sigma_I$ although σ^2_I is very messy to compute, all the terms are found in the coefficients in the equations to find the weights.

The correlated response for any trait c can be computed as usual as

$\triangle G_c = \dfrac{\text{Cov}(G_{\alpha c}, I_\alpha)}{\sigma_I} D$. Again $\text{Cov}(G_{\alpha c}, I_\alpha)$ is very messy but can be computed:

$$\text{Cov}(G_{\alpha c}, I_\alpha) = \beta_1 \text{Cov}(G_{\alpha c}, X_1) + \beta_2 \text{Cov}(G_{\alpha c}, X_2) + \ldots + \beta_N \text{Cov}(G_{\alpha c}, X_N)$$

where $\text{Cov}(G_{\alpha c}, X_k) = \text{Cov}(G_{\alpha c}, X_{ij}) =$

$a_{i\alpha} \sigma_{G_c G_j}$ and (if c = j) = $a_{i\alpha} \sigma^2_{G_c}$.

Approximate procedure

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. 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 I directly.

The approximate procedure is the same as using records of relatives for only the trait being indexed. Then the indexes for each trait (based only on records for that trait) are weighted by their economic values as before.

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

The following example with two traits measured on the animal and on 50 paternal half sibs (p.h.s.) will illustrate the exact and approximate procedures and will demonstrate how to compare the expected selection response from both.

The following problem illustrates use of standardized variables when three traits are measured on the animal being evaluated and when the three traits have economic value. A part of the problem also illustrates the consequences of assuming the genetic and phenotypic correlations are all zero.

Problem

Given: $v_1' = 3$, $v_2' = 2$, $v_3' = 1$ (the relative economic values of standard deviation units)

$\sigma_{X_1} = 6$, $\sigma_{X_2} = 5$, $\sigma_{X_3} = 4$, $h_1^2 = .7$, $h_2^2 = .8$, $h_3^2 = .9$

$r_{p_{12}} = .1$, $r_{p_{13}} = .2$, $r_{p_{23}} = .3$, $r_{g_{12}} = .6$, $r_{g_{13}} = .5$, $r_{g_{23}} = .4$

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

$$b_{11}' + .1\, b_{12}' + .2\, b_{13}' = .7$$

$$.1\, b_{11}' + b_{12}' + .3\, b_{13}' = .6\sqrt{(.7)(.8)}$$

$$.2\, b_{11}' + .3\, b_{12}' + b_{13}' = .5\sqrt{(.7)(.9)} ,$$

$$I_1' = .632\, Y_1 + .335\, Y_2 + .170\, Y_3$$

For I_2' the RHS's become $.6\sqrt{(.8)(.7)}$, .8, and $.4\sqrt{(.8)(.9)}$ so that

$$I_2' = .366\, Y_1 + .751\, Y_2 + .041\, Y_3$$

For I_3' the RHS's are $.5\sqrt{(.9)(.7)}$, $.4\sqrt{(.9)(.8)}$, and .9 and

$$I_3' = .223\, Y_1 + .066\, Y_2 + .835\, Y_3 .$$

Or, to find I' directly solve

$$\beta_1' + .1\, \beta_2' + .2\, \beta_3' = 3(.7) + 2(.6)\sqrt{(.7)(.8)} + (1)(.5)\sqrt{(.7)(.9)}$$

$$.1\, \beta_1' + \beta_2' + .3\, \beta_3' = 3(.6)\sqrt{(.8)(.7)} + 2(.8) + (1)(.4)\sqrt{(.8)(.9)}$$

$$.2\, \beta_1' + .3\, \beta_2' + \beta_3' = 3(.5)\sqrt{(.9)(.7)} + 2(.4)\sqrt{(.9)(.8)} + (1)(.9)$$

and $I' = 2.85\, Y_1 + 2.57\, Y_2 + 1.43\, Y_3 = I$ and $\sigma_{I'} = 4.7$.

The correlated response in trait 2 is

$$\triangle_c G_2' = \frac{(2.85)(.6)\sqrt{(.8)(.7)} + (2.57)(.8) + (1.43)(.4)\sqrt{(.8)(.9)}}{4.7} D$$

$$= .813\ D$$

and $\triangle_c G_2 = 5(.813)D = 4.065\ D$.

Now assume r_g's and r_p's = 0 as for the approximate index.

The overall equations reduce to $\beta_1' = 3(.7)$, $\beta_2' = 2(.8)$, and $\beta_3' = (1)(.9)$ so that $I' = 2.1\ Y_1 + 1.6\ Y_2 + .9\ Y_3$, $\sigma^2_{I'} = 7.78$ and $\sigma_{I'} = 2.79$, $Cov(G_2', I') = 1.28$ and $\triangle_c G_2 = 2.29\ D$ 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 $\sigma^2_{I'} = 7.78 + \underline{2.29}$, $\sigma_{I'} = 3.17$; $Cov(G_2', I') = 2.53$ and

$\triangle_c G_2 = (5)\frac{2.53}{3.17} D = 3.99\ D$ as compared to 4.06 D using the best index and to the 2.29D expected by assuming zero correlations.

Another standardization

Some research reports have used another standardization procedure which gives all 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:

$$Y = \frac{X - \mu_X}{\sigma_{G_X}} = \frac{G_X}{\sigma_{G_X}} + \frac{E_X}{\sigma_{G_X}} .$$

$$V\left(\frac{G_X}{\sigma_{G_X}}\right) = 1 \quad \text{and} \quad V\left(\frac{E_X}{\sigma_{G_X}}\right) = \frac{1-h^2}{h^2} \quad \text{so that}$$

$$V(Y) = \frac{1}{h^2} .$$

For standardized records on two traits the phenotypic covariance is

$$Cov(Y_1, Y_2) = Cov\left(\frac{X_1}{\sigma_{G_{Y_1}}}, \frac{X_2}{\sigma_{G_{Y_2}}}\right) = \frac{\sigma_{X_1X_2}}{\sigma_{G_{X_1}}\sigma_{G_{X_2}}} = r_{p_{12}} \frac{1}{\sqrt{h_1^2 h_2^2}} .$$

The genetic covariance is $Cov(G_{Y_1}, G_{Y_2}) = Gov\left(\frac{G_{X_1}}{\sigma_{G_{X_1}}}, \frac{G_{X_2}}{\sigma_{G_{X_2}}}\right) = \frac{\sigma_{G_{X_1}G_{X_2}}}{\sigma_{G_{X_1}}\sigma_{G_{X_2}}} = r_{g_{12}}$.

If records are standardized in this way the equations which determine the proper weights (b''_{mi}, i=1, ..., N) when selecting for trait m using standardized records on the animal to be evaluated, $I''_m = b''_{m1} Y_1 + \ldots + b''_{mN} Y_N$, are

$$\begin{aligned}
\frac{1}{h_1^2} b''_{m1} + \frac{1}{\sqrt{h_1^2 h_2^2}} r_{p_{12}} b''_{m2} + \ldots + \frac{1}{\sqrt{h_1^2 h_N^2}} r_{p_{1N}} b''_{mN} &= r_{g_{m1}} \\
\vdots & \quad \text{(when } m = 1, \ r_{g_{mi}} = 1\text{)} \\
\frac{1}{\sqrt{h_1^2 h_N^2}} r_{p_{1N}} b''_{m1} + \frac{1}{\sqrt{h_2^2 h_N^2}} r_{p_{2N}} b''_{m2} + \ldots + \frac{1}{h_N^2} b''_{mN} &= r_{g_{mN}}
\end{aligned}$$

The extension of this to $T = \sum_i v''_m G_m$ is straightforward. The economic values are given in terms of value per genetic standard deviation, v''_m . The index for trait m in standardized form can be converted back to non-standardized form as

$I_m = \sigma_{G_m} I''_m$. Similarly $\hat{T} = I = v_1 I_1 + \ldots + v_M I_M$.

Summary XII

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, very difficult.

One method of analysis is to simply assign a single score to each birth. Two 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 economic value is implied. In both cases the usual selection index procedure can be used if the appropriate heritability is known.

The best procedure, however, is the selection index procedure with each category being considered as a separate trait scored as zero or one. There will, however, be covariances among the categories. Categorical data have a multinomial distribution. If there are only two categories the distribution is the usual binomial one.

The phenotypic variances of and covariances among the categories are determined by the probabilities of being scored in each category. These fractions 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:

$$\begin{pmatrix} \sigma^2_{y_1} & \sigma_{y_1y_2} & \sigma_{y_1y_3} \\ \sigma_{y_1y_2} & \sigma^2_{y_2} & \sigma_{y_2y_3} \\ \sigma_{y_1y_3} & \sigma_{y_2y_3} & \sigma^2_{y_3} \end{pmatrix} = \begin{pmatrix} \pi_1(1-\pi_1) & -\pi_1\pi_2 & -\pi_1\pi_3 \\ -\pi_1\pi_2 & \pi_2(1-\pi_2) & -\pi_2\pi_3 \\ -\pi_1\pi_3 & -\pi_2\pi_3 & \pi_3(1-\pi_3) \end{pmatrix}.$$

There are some distinct properties to consider when dealing with multinomial data. 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. The sum of the variances and covariances in any row (or column) is zero.

$$\begin{pmatrix} \sigma^2_{g_1} & \sigma_{g_1g_2} & \sigma_{g_1g_3} \\ \sigma_{g_1g_2} & \sigma^2_{g_2} & \sigma_{g_2g_3} \\ \sigma_{g_1g_3} & \sigma_{g_2g_3} & \sigma^2_{g_3} \end{pmatrix}$$

Such a property results in what is known as a lack of independence. Such variance-covariance matrices are singular. The practical result is that instead of using all the traits in predicting the value for any one trait as is usual for evaluation using multiple traits, all traits except one are used as will be illustrated.

The selection indexes predict differences from the means as follows:

$$I_1 = \hat{g}_1 - \pi_1 = b_{11}(X_1-\pi_1) + b_{12}(X_2-\pi_2) + b_{13}(X_3-\pi_3)$$
$$I_2 = \hat{g}_2 - \pi_2 = b_{21}(X_1-\pi_1) + b_{22}(X_2-\pi_2) + b_{23}(X_3-\pi_3)$$
$$I_3 = \hat{g}_3 - \pi_3 = b_{31}(X_1-\pi_1) + b_{32}(X_2-\pi_2) + b_{33}(X_3-\pi_3)$$

Then 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 $I_1 + I_2 + I_3 = 0$. Remember that 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, $$\sigma^2_{X_i} = \frac{\sigma^2_{y_i} + (p-1)(\frac{1}{4}\sigma^2_{g_i})}{p}$$ and

$$\sigma_{X_i X_j} = \frac{\sigma_{y_i y_j} + (p-1)(\frac{1}{4}\sigma_{g_i g_j})}{p} .$$

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 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:

For g_1 : $I_1 = .20(X_1-\pi_1) + .0(X_2-\pi_2) + 0(X_3-\pi_3)$

For g_2 : $I_2 = .14(X_1-\pi_1) + .433(X_2-\pi_2) + 0(X_3-\pi_3)$

For g_3 : $I_3 = -.34(X_1-\pi_1) - .433(X_2-\pi_2) + 0(X_3-\pi_3)$

and $\hat{g}_1 = I_1 + \pi_1$, $\hat{g}_2 = I_2 + \pi_2$, $\hat{g}_3 = I_3 + \pi_3$.

When the animal is scored in category 1, $X_1 = 1$, $X_2 = 0$, and $X_3 = 0$ so that

$\hat{g}_1 = .20(1-.5) + 0(0-.3) + 0(0-.2) + .5 = .60$,

$\hat{g}_2 = .14(1-.5) + .433(0-.3) + 0(0-.2) + .3 = .24$, and

$\hat{g}_3 = -.34(1-.5) - .433(0-.3) + 0(0-.2) + .2 = .16$.

When the animal is scored in category 2, $X_1 = 0$, $X_2 = 1$, and $X_3 = 0$:

$$\hat{g}_1 = .20(0-.5) + 0(1-.3) + 0(0-.2) + .5 = .40,$$

$$\hat{g}_2 = .14(0-.5) + .433(1-.3) + 0(0-.2) + .3 = .5331, \text{ and}$$

$$\hat{g}_3 = -.34(0-.5) - .433(1-.3) + 0(0-.2) + .2 = .0669 .$$

When the animal is scored in category 3, $X_1 = 0$, $X_2 = 0$, and $X_3 = 1$:

$$\hat{g}_1 = .20(0-.5) + 0(0-.3) + 0(1-.2) + .5 = .40,$$

$$\hat{g}_2 = .14(0-.5) + .433(0-.3) + 0(1-.2) + .3 = .10, \text{ and}$$

$$\hat{g}_3 = -.34(0-.5) - .433(0-.3) + 0(1-.2) + .2 = .50 .$$

If any other two equations had been used, e.g. X_2 and X_3 with $b_1 = 0$, the evaluation would have been exactly the same. Note that the appropriate RHS's to predict g_1, g_2, and g_3 would have been

$$\begin{pmatrix} -.03 & .07 & -.04 \\ ; & ; & \\ -.02 & -.04 & .06 \end{pmatrix} .$$

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 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 to the frequencies which 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: $60(.40) - 5(.5331) - 100(.0669) = -45.5655$.

Summary XIII

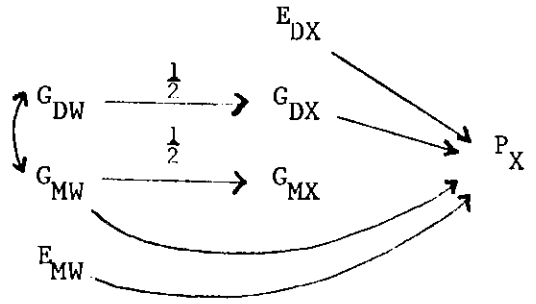
SELECTION FOR EMBEDDED TRAITS

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. Three examples of embedded traits will be discussed in this section: the maternal effects model, the grandmaternal effects model and the fetal 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 half of her genes that the offspring has obtained. In turn, part of the maternal effect may be genetic and part may be environmental. (See Willham, Biometrics, 1963 for the 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 genetic effect associated with the genotype of X, G_{MW} is the genetic maternal effect on X, E_{MW} is the environmental maternal effect on X, G_{MX} is the genetic maternal ability of X which is not measured, and G_{DW} is the genetic effect associated with the dam of X, W. Note that $G_{MW} + E_{MW} + E_{DX} = E_X$ and $G_{DX} = G_X$ of the usual model $P_X = G_X + E_X$.

One can consider these to be two traits (a direct trait, D, and an indirect maternal trait, M) which may be correlated. Trait M is measured one generation later than the direct effect D.

The genetic covariances between relatives with maternal effects considered

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.

By the rules for the covariance of linear functions we can find the genetic covariance between X and Y. Assuming any environmental covariances are zero, we find:

$$COV(P_X, P_Y) = COV(G_{DX}, G_{DY}) + COV(G_{DX}, G_{MZ}) + COV(G_{MW}, G_{DY}) + COV(G_{MW}, G_{MZ}) .$$

In terms of genetic variance and covariance components we have:

$$COV(G_{DX}, G_{DY}) = a_{XY} \frac{\sigma^2_{10}}{D} + a^2_{XY} \frac{\sigma^2_{20}}{D} + d_{XY} \frac{\sigma^2_{01}}{D} + \ldots ,$$

$$COV(G_{MW}, G_{MZ}) = a_{WZ} \frac{\sigma^2_{10}}{M} + a^2_{WZ} \frac{\sigma^2_{20}}{M} + d_{WZ} \frac{\sigma^2_{01}}{M} + \ldots ,$$

$$COV(G_{DX}, G_{MZ}) = a_{XZ} \frac{\sigma_{10},\frac{10}{M}}{D} + a^2_{XZ} \frac{\sigma_{20},\frac{20}{M}}{D} + d_{XZ} \frac{\sigma_{01},\frac{01}{M}}{D} + \ldots , \text{ and}$$

$$COV(G_{MW}, G_{DY}) = a_{WY} \frac{\sigma_{10},\frac{10}{M}}{D} + a^2_{WY} \frac{\sigma_{20},\frac{20}{M}}{D} + d_{WY} \frac{\sigma_{01},\frac{01}{M}}{D} + \ldots .$$

The a's and d's are the usual additive and dominance relationships. The genetic variances are labelled with the trait, i.e., (σ^2_{10}/D) is the additive genetic variance of the direct trait, D.

The genetic covariances are labelled with both traits, i.e., $\frac{\sigma_{10},\frac{10}{M}}{D}$ is the covariance between the additive genetic effects for trait D and trait M.

If only additive genetic effects are considered,

$$COV(P_X P_Y) = a_{XY} \frac{\sigma^2_{10}}{D} + a_{WZ} \frac{\sigma^2_{10}}{M} + (a_{XZ} + a_{WY}) \frac{\sigma_{10},\frac{10}{M}}{D} =$$

$$a_{XY} \sigma^2_{G_D} + a_{WZ} \sigma^2_{G_M} + (a_{XZ} + a_{WY}) \sigma_{G_D G_M} \text{ in a simpler notation.}$$

Example of genetic covariances between relatives considering only additive effects

Animal with itself: This covariance will be the genetic variance plus environmental variance.

$X = Y$, $W = Z$, and $a_{XY} = 1$, $a_{WZ} = 1$, $a_{XW} = \frac{1}{2}$, and $a_{YZ} = \frac{1}{2}$.

$$COV(P_X P_X) = \sigma^2_{G_D} + \sigma^2_{G_M} + (\tfrac{1}{2} + \tfrac{1}{2})\sigma_{G_D G_M} + \sigma^2_{E_M} + \sigma^2_{E_D} = \sigma^2_P .$$

In terms of the P = G + E model $\sigma^2_G = \sigma^2_{G_D}$ and $\sigma^2_E = \sigma^2_{G_M} + \sigma_{G_D G_M} + \sigma^2_{E_M} + \sigma^2_{E_D}$.

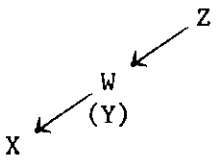
Dam-progeny covariance considering only additive genetic effects:

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:

$$P_X = G_{DX} + G_{\underset{(Y)}{MW}} \quad \text{and} \quad P_{\underset{(W)}{Y}} = G_{\underset{(W)}{DY}} + G_{MZ} .$$

$$a_{XY} = 1/2, \quad a_{\underset{(Y)}{WZ}} = 1/2, \quad a_{XZ} = 1/4, \quad a_{WY} = 1 .$$

Then $COV(P_X,P_Y) = 1/2\ \sigma^2_{G_D} + 1/2\ \sigma^2_{G_M} + (1/4 + 1)\ \sigma_{G_DG_M}$.

Note that the covariance between relatives may contain a genetic covariance between the direct and maternal traits. This covariance can be negative and thus mask the additive genetic variances for the direct and maternal traits.

Note that the additive genetic correlation between D and M is $r_{g_{D,M}} = \dfrac{\sigma_{G_DG_M}}{\sqrt{\sigma^2_{G_D}\sigma^2_{G_M}}}$. Since the maximum absolute value of r_g is 1,

$$\sqrt{\sigma^2_{G_D}\sigma^2_{G_M}} \geq \left|\sigma_{G_D}\sigma_{G_M}\right| .$$

Thus, there is a possibility of obtaining a negative estimate of the offspring-dam covariance if the negative value of $\sigma_{G_DG_M}$ is large enough.

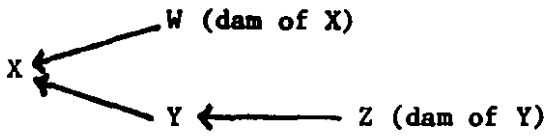
If maternal effects are important doubling the offspring on parent regression can give a biased estimate of heritability of the direct trait;

i.e., $$h^2 = \frac{2[1/2\ \sigma^2_{G_D} + 1/2\ \sigma^2_{G_M} + 1\ \frac{1}{4}\ \sigma_{G_DG_M}]}{\sigma^2_P} = \frac{\sigma^2_{G_D}}{\sigma^2_P} + \frac{\sigma^2_{G_M} + 2\ \frac{1}{2}\ \sigma_{G_DG_M}}{\sigma^2_P}$$

Of course there are also the other possible genetic causes for bias in this estimate due to σ^2_{20} etc.

Sire-progeny covariance:

X is progeny of dam W and sire Y which has dam Z.



$a_{XY} = \frac{1}{2}$, $a_{WZ} = 0$, $a_{XZ} = \frac{1}{4}$, $a_{YW} = 0$

then $Cov(P_X P_Y) = \frac{1}{2}\sigma^2_{G_D} + \frac{1}{4}\sigma_{G_D G_M}$ which is quite different from the offspring-dam covariance.

Practice Problems

The following problems illustrate some concepts of covariance among relatives when maternal traits are important.

1. Estimate $\sigma^2_{G_D}$, $\sigma^2_{G_M}$, $\sigma_{G_D G_M}$

   Given: Covariance between paternal half sibs = 20

   Covariance between full sibs = 30

   Covariance between offspring & sire = 30

2. Given: $\sigma^2_{G_D} = 80$, $\sigma^2_{G_M} = 40$, $\sigma_{G_D G_M} = -20$

   Show all calculations (steps) in computing the covariances between:

   a) offspring and dam

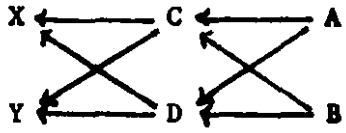
   b) offspring and sire

   c) full sibs

   d) maternal half sibs

   e) paternal half sibs

   f) X and Y in diagram



Solutions:

1. Cov (p h s) $= 1/4\ \sigma^2_{G_D} + 0\ \sigma^2_{G_M} + 0\ \sigma_{G_D G_M} = 20$ [1]

   Cov (full sibs) $= 1/2\ \sigma^2_{G_D} + \sigma^2_{G_M} + (1)\sigma_{G_D G_M} = 30$ [2]

   Cov (offs, sire) $= 1/2\ \sigma^2_{G_D} + 0\ \sigma^2_{G_M} + 1/4\sigma_{G_D G_M} = 30$ [3]

   From [1]: $\sigma^2_{G_D} = 4(20) = 80$

   From [1] and [3]: $1/2(80) + 1/4\ \sigma_{G_D G_M} = 30$; $\sigma_{G_D G_M} = -40$

   From [2]: $1/2(80) + \sigma^2_{G_M} + (-40) = 30$; $\sigma^2_{G_M} = 30$

2. a) $X \leftarrow Y \leftarrow Z$ (W) $\quad a_{XY} = 1/2 \quad a_{WZ} = 1/2 \quad a_{XZ} = 1/4 \quad a_{YW} = 1$

   Cov (offspring-dam) = 1/2(80) + 1/2(40) + (1/4 + 1)(-20) = 35

   b) W → X, Y ← Z; Y = sire $\quad a_{XY} = 1/2 \quad a_{WZ} = 0 \quad a_{XZ} = 1/4 \quad a_{YW} = 0$

   Cov (offspring-sire) = 1/2(80) + 0 + (1/4 + 0)(-20) = 35

   c) X ← sire, Y ← dam (W & Z) $\quad a_{XY} = 1/2 \quad a_{WZ} = 1 \quad a_{XZ} = 1/2 \quad a_{YW} = 1/2$

   Cov (full-sibs) = 1/2(80) + (1)(40) + (1/2 + 1/2)(-20) = 60

   d) X, Y ← dam (W & Z) $\quad a_{XY} = 1/4 \quad a_{WZ} = 1 \quad a_{XZ} = 1/2 \quad a_{YW} = 1/2$

   Cov (mat. half sibs) = 1/4(80) + (1)(40) + (1/2 + 1/2)(-20) = 40

   e) X ← W (dam of X); Sire; Y ← Z (dam of Y) $\quad a_{XY} = 1/4 \quad a_{WZ} = 0 \quad a_{XZ} = 0 \quad a_{YW} = 0$

   Cov (pat. half sibs) = 1/4(80) + 0 + (0 + 0) = 20

   f) X ← sire ← A; C; dam; Y ← D ← B; W&Z $\quad a_{XY} = 3/4 \quad a_{WZ} = 1 \quad a_{XZ} = 3/4 \quad a_{yw} = 3/4$

   Cov (X and Y) = 3/4(80) + 1(40) + (3/4 + 3/4)(-20) = 70

The following table gives the additive relationships that are coefficients of $\sigma^2_{G_D}$, $\sigma^2_{G_M}$, and $\sigma_{G_D G_M}$ for 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 $\text{Cov}(P_X, P_Y) = a_{XY}\sigma^2_{G_D} + a_{WZ}\sigma^2_{G_M} + (a_{XZ} + a_{YW})\sigma_{G_D G_M}$

P_X, P_Y	a_{XY}	a_{WZ}	a_{XZ}	a_{YW}
P_X, P_X (with self)	1	1	1/2	1/2
Progeny, dam	1/2	1/2	1/4	1
Progeny, sire	1/2	0	1/4	0
Full sibs	1/2	1	1/2	1/2
Maternal sibs	1/4	1	1/2	1/2
Paternal sibs	1/4	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.

RHS's: $\sigma_{X_i G_{D\alpha}} = a_{i\alpha}\sigma^2_{G_D} + a_{W_i\alpha}\sigma_{G_D G_M}$

$= E(X_i, G_{D\alpha}) = E[(G_{D_i} + G_{M_{W_i}} + E_{D_i} + E_{M_{W_i}})(G_{D\alpha})]$,

where W_i is the dam of i and α is the animal being evaluated.

Diagonal coefficients: The $(p_i - 1)$ coefficient will be expanded by the maternal variance and maternal-direct covariance.

$\sigma^2_{X_i} = \{\sigma^2_P + (p_i - 1)[(a_{ii'}\sigma^2_{G_D} + a_{W_iW_{i'}}\sigma^2_{G_M} + (a_{iW_{i'}} + a_{i'W_i})\sigma_{G_D G_M}]\}/p$ where $a_{ii'}$ is relationship among members of the group, $a_{W_iW_{i'}}$ is relationship among dams of group, $a_{iW_{i'}}$ is relationship of animal in group to another's dam.

Off-diagonal coefficients: These will be expanded similarly.

$\sigma_{X_iX_j} = a_{ij}\sigma^2_{G_D} + a_{W_iW_j}\sigma^2_{G_M} + (a_{iW_j} + a_{jW_i})\sigma_{G_D G_M}$.

Selection for the maternal trait.

RHS's: $\sigma_{X_i G_{M\alpha}} = a_{i\alpha}\, \sigma_{G_D G_M} + a_{W_i\alpha}\, \sigma^2_{G_M}$

Diagonal and off-diagonal coefficients will be as in selection for the direct trait.

Correlated response when selecting for G_D or G_M.

If selection is for $G_{D\alpha}$ 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 $\triangle G_D = \sigma_{I_D} D$ (the selection differential D is different from the subscript D which refers to the direct trait), where $I_D = b_{1D}X_1 + \ldots + b_{ND}X_N$.

The response in G_D can also be computed as usual as:

$$\triangle G_D = \frac{\text{Cov}\,(G_{D\alpha}, I_D)}{\sigma_{I_D}} D = \frac{\sum_{i=1}^{N} b_{iD}\, \text{Cov}\,(G_{D\alpha}, X_i)}{\sigma_{I_D}} D \quad \text{where}$$

$$\text{Cov}\,(G_{D\alpha}, X_i) = a_{i\alpha}\, \sigma^2_{G_D} + a_{W_i\alpha}\, \sigma_{G_D G_M} .$$

Similarly the correlated response in G_M can be predicted as

$$\triangle G_M = \frac{\text{Cov}\,(G_{M\alpha}, I_D)}{\sigma_{I_D}} D = \frac{\sum_{i=1}^{N} b_{iD}\, \text{Cov}\,(G_{M\alpha}, X_i)}{\sigma_{I_D}} D \quad \text{where}$$

$$\text{Cov}\,(G_{M\alpha}, X_i) = a_{i\alpha}\, \sigma_{G_D G_M} + a_{W_i\alpha}\sigma^2_{G_M} .$$

If selection is for $G_{M\alpha}$ by $I_M = \sum_{i=1}^{N} b_{iM} X_i$

$$\triangle G_D = \frac{\text{Cov}\,(G_{D\alpha}, I_M)}{\sigma_{I_M}} D = \frac{\sum_{i=1}^{N} b_{iM}\, \text{Cov}\,(G_{D\alpha} X_i)}{\sigma_{I_M}} D \quad \text{and}$$

$$\triangle G_M = \sigma_{I_M} D \quad \text{or} \quad \triangle G_M = \frac{\sum_{i=1}^{N} b_{iM}\, \text{Cov}\,(G_{M\alpha} X_i)}{\sigma_{I_M}} D .$$

The following example illustrates computations for these concepts and also shows how to compute the effect of bias in heritability estimates if maternal effects are ignored.

Example of selection for direct genetic effects

Given: $\sigma^2_{G_D} = 80$, $\sigma^2_{G_M} = 40$, $\sigma_{G_D G_M} = 40$, $\sigma^2_P = 500$

Also suppose heritability is estimated in the usual way by twice the regression of offspring on dam record.

1. a) Use this biased estimate of heritability (genetic variance) to find the usual weights for indexing genetic value from the animal's own record, X_1, and the sire's record, X_2.
   b) What is the expected progress by the usual procedure of calculating genetic gain?
   c) Use the incorrect index found in (1 a) 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

$\sigma^2_{G_D} = 80$, $\sigma^2_{G_M} = 40$, $\sigma_{G_D G_M} = 40$, $\sigma^2_P = 500$, X_1 = own record, X_2 = sire's record.

Heritability from twice regression of offspring on dam record

$$\hat{h}^2 = 2 \text{ Cov (offspring, dam)}/\sigma^2_P = 2\left[\tfrac{1}{2}(80) + \tfrac{1}{2}(40) + (1\tfrac{1}{4})(40)\right]/500 = .44$$

1. a) $b_1 + (\frac{1}{2})(.44)\, b_2 = .44$

$$.22\, b_1 + b_2 = \tfrac{1}{2}(.44)$$

$$I = .4115\, X_1 + .1295\, X_2 = \hat{G}_\alpha$$

$\sigma_I = 10.24$; would be the apparent standard deviation of index. The actual σ_I is 9.92 since $\sigma_{X_1 X_2} = 50$ while $\hat{h}^2 = .44$ implies $\sigma_{X_1 X_2} = 110$.

b) $\triangle$ $G = 10.24\ D$ would be the usual prediction based on $h^2 = .44$.

c) $\triangle_c$ $G_D = \dfrac{b_1\ Cov(G_{D\alpha},X_1) + b_2\ Cov(G_{D\alpha},X_2)}{\sigma_I} D =$

$$\frac{.4115[(1)(80)+(\frac{1}{2})(40)]+.1295[(\frac{1}{2})(80)+(\frac{1}{4})(40)]}{9.92} D = 4.80\ D$$

$\triangle_c$ $G_M = \dfrac{b_1\ Cov(G_{M\alpha},X_1) + b_2\ Cov(G_{M\alpha},X_2)}{\sigma_I} D =$

$$\frac{.4115[(1)(40)+(\frac{1}{2})(40)]+.1295[(\frac{1}{2})(40)+(\frac{1}{4})(40)]}{9.92} D = 2.88\ D$$

2. a) $500\ b_1 + [\frac{1}{2}(80)+(\frac{1}{4})(40)]\ b_2 = (1)\ 80 + (\frac{1}{2})(40)$

$50\ b_1 + 500\ b_2 = \frac{1}{2}(80) + (\frac{1}{4})(40)$

$I = .1919\ X_1 + .0808\ X_2 = \hat{G}_{D\alpha}$

$\sigma_I = 4.82$

b) $\triangle_c$ $G_D = \dfrac{b_1\ Cov(G_{D\alpha},X_1) + b_2\ Cov(G_{D\alpha},X_2)}{\sigma_I} D = \dfrac{.1919(100)+.0808(50)}{4.82} D = 4.82\ D$

$\triangle_c$ $G_M = \dfrac{b_1\ Cov(G_{M\alpha},X_1) + b_2\ Cov(G_{M\alpha},X_2)}{\sigma_I} D = \dfrac{.1919(60)+.0808(30)}{4.82} D = 2.89\ D$

3. $\sigma^2_{G_D} = 80 \quad \sigma^2_{G_M} = 40 \quad \sigma_{G_D G_M} = -40, \quad \sigma^2_P = 500$

Now heritability from twice offspring on parent regression

$\hat{h}^2 = 2\ Cov(offspring,dam)/\sigma^2_P = 2[\frac{1}{2}(80)+\frac{1}{2}(40)+1\frac{1}{4}(-40)]/500 = .04$

(1.a) $b_1 + \frac{1}{2}(.04)\ b_2 = .04$

$.02\ b_1 + b_2 = \frac{1}{2}(.04)$

$I = .0396\ X_1 + .0192\ X_2 = \hat{G}_\alpha$; $\sigma_I = .9918$ is apparent σ_I. The actual

$\sigma_I = 1.007$.

(1.b) $\triangle$ $G = .9918\ D$ usual prediction with $h = .04$

(1.c) $\triangle_c$ $G_D = \dfrac{b_1\ Cov(G_{D\alpha},X_1) + b_2\ Cov(G_{D\alpha},X_2)}{\sigma_I} D =$

$$\frac{.0396[(1(80)-\frac{1}{2}(40)]+.0192[\frac{1}{2}(80)-\frac{1}{4}(40)]}{1.007} D = 2.931\ D$$

$$\triangle_c\ G_M = \frac{b_1\ \mathrm{Cov}\ G_{M\alpha},X_1)+ b_2\ \mathrm{Cov}(G_{M\alpha},X_2)}{\sigma_I} D =$$

$$\frac{.0396[(1)(-40)+ \frac{1}{2}(40)]+.0192[\frac{1}{2}(-40)+ \frac{1}{4}(40)]}{1.007} D = -.977\ D$$

(2.a) $500\ b_1 + [\frac{1}{2}(80) + \frac{1}{4}(-40)]\ b_2 = (1)(80)+(\frac{1}{2})(-40)$

$30\ b_1 + 500\ b_2 = (\frac{1}{2})(80)+ \frac{1}{4}(-40)$

$I = .1168\ X_1 + .0530\ X_2 = \hat{G}_{D\alpha}$

$\sigma_I = 2.932$

(2.b) $$\triangle_c\ G_D = \frac{b_1\ \mathrm{Cov}(G_{D\alpha},X_1)+ b_2\ \mathrm{Cov}(G_{D\alpha},X_2)}{\sigma_I} D =$$

$$\frac{.1168(60) + .0530(30)}{2.932} D = 2.932\ D$$

$$\triangle_c\ G_M = \frac{b_1\ \mathrm{Cov}(G_{M\alpha},X_1)+ b_2\ \mathrm{Cov}(G_{M\alpha},X_2)}{\sigma_I} D =$$

$$\frac{.1168(-20) + .0530(-10)}{2.932} D = -.977\ D$$

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 values are not necessarily the same because, although the gross price is the same for the total product, the cost of production may be greatly 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 since 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 only records on two relatives, X_1 and X_2. The index will be $I_\alpha = b_1 X_1 + b_2 X_2$ which estimates

$$T_\alpha = v_D G_{D\alpha} + v_M G_{M\alpha} .$$

The general equations which determine the b's are

$$b_1 \sigma^2_{X_1} + b_2 \sigma_{X_1X_2} = \sigma_{X_1T_\alpha}$$

$$b_1 \sigma_{X_1X_2} + b_2 \sigma^2_{X_2} = \sigma_{X_2T_\alpha} .$$

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

$$\sigma_{X_iT_\alpha} = v_D \sigma_{X_iG_{D\alpha}} + v_M \sigma_{X_iG_{M\alpha}}$$

$$= v_D (a_{i\alpha}\sigma^2_{G_D} + a_{W_i\alpha}\sigma_{G_DG_M}) + v_M (a_{i\alpha}\sigma_{G_DG_M} + a_{W_i\alpha}\sigma^2_{G_M})$$

Then $I_\alpha = b_1 X_1 + b_2 X_2$. An alternative procedure would be to index for $G_{D\alpha}$ and $G_{M\alpha}$ separately then weight 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 $I_{M\alpha} = \hat{G}_{M\alpha}$ as in the preceding sections.

The expected response by selection can be computed as before

$$\triangle T = v_D \triangle G_D + v_M \triangle G_M \quad \text{where}$$

$$\triangle G_D = \frac{\text{COV}(G_{D\alpha}, I)}{\sigma_I} D \quad \text{and} \quad \triangle G_M = \frac{\text{COV}(G_{M\alpha}, I)}{\sigma_I} D .$$

The following examples 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 in selecting for G_D and G_M simultaneously

Given: $\sigma^2_{G_D} = 80$ $\sigma^2_{G_M} = 40$ $\sigma_{G_D G_M} = 40$, $\sigma^2_P = 500$

X_1 = record on the sire, X_2 = record on the dam.

1. If $v_D = 4$ and $v_M = 1$ can we select 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

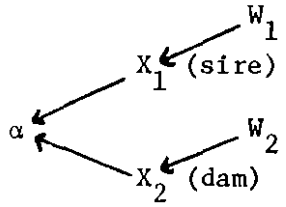
   What is the index for selecting for $T_\alpha = (4)\, G_{D\alpha} + (1)\, G_{M\alpha}$?

   b) What is the expected correlated response in G_D and G_M ?

3. Repeat (2 a and b) when $\sigma_{G_D G_M} = -40$.

Solutions

1. X_1 = record on sire X_2 = record on dam



   $(a_{1\alpha}\ ,\ a_{W_1\alpha}) = (1/2\ ,\ 1/4)$;

   $(a_{2\alpha}\ ,\ a_{W_2\alpha}) = (1/2\ ,\ 1/4)$. These are

   proportional so selection cannot be for $v_D G_D + v_M G_M$.

   For example:

   If $v_D = 4$ $v_M = 1$ $I = .46\, X_1 + .46\, X_2$, $\sigma_I = 14.55$

   $$\triangle G_D = \frac{.46[50] + .46[50]}{14.55} D = 3.16\, D$$

   $$\triangle G_M = \frac{.46[30] + .46[30]}{14.55} D = 1.90\, D$$

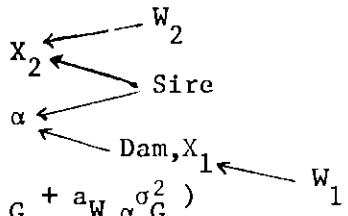
   If $v_D = 1$ $v_M = 4$ $I = .34\, X_1 + .34\, X_2$, $\sigma_I = 10.75$ and

   $$\triangle G_D = \frac{.34[50] + .34[50]}{10.75} D = 3.16\, D$$

   $$\triangle G_M = \frac{.34[30] + .34[30]}{10.75} D = 1.90\, D$$ as before for $v_D = 4$ and $v_M = 1$.

2. X_1 = record on dam X_2 = record on p.h.s.

$v_D = 4$ $v_M = 1$



a) $500b_1 + 0\ b_2 = v_D(a_{1\alpha}\sigma^2_{G_D} + a_{W_1\alpha}\sigma_{G_M G_D}) + v_M(a_{1\alpha}\sigma_{G_M G_D} + a_{W_1\alpha}\sigma^2_{G_M})$

$0\ b_1 + 500b_2 = v_D(a_{2\alpha}\sigma^2_{G_D} + a_{W_2\alpha}\sigma_{G_M G_D}) + v_M(a_{2\alpha}\sigma_{G_M G_D} + a_{W_2\alpha}\sigma^2_{G_M})$

$500b_1 = 4[\frac{1}{2}(80) + \frac{1}{4}(40)] + 1[\frac{1}{2}(40) + \frac{1}{4}(40)] = 230$

$500\ b_2 = 4[\frac{1}{4}(80) + 0\ ] + 1[\frac{1}{4}(40) + 0\ ] = 90$

$I = .46\ X_1 + .18\ X_2$

$\sigma_I = 11.05$

b) $\hat{G}_D = \frac{COV(G_{D\alpha},\ I)}{\sigma_I} D = \frac{.46[\frac{1}{2}(80) + \frac{1}{4}(40)] + .18[\frac{1}{4}(80) + 0]}{11.05} D = 2.407\ D$

$\hat{G}_M = \frac{COV(G_{M\alpha},\ I)}{\sigma_I} D = \frac{.46[\frac{1}{2}(40) + \frac{1}{4}(40)] + .18[\frac{1}{4}(40) + 0]}{11.05} D = 1.412\ D$

3. $\sigma_{G_D G_M} = -40$

(2.a) $500\ b_1 + 0\ b_2 = 4[\frac{1}{2}(80) + \frac{1}{4}(-40)] + 1[\frac{1}{2}(-40) + \frac{1}{4}(40)] = 110$

$0\ b_1 + 500\ b_2 = 4[\frac{1}{4}(80) + 0\ ] + 1[\frac{1}{4}(-40) + 0\ ] = 70$

$I = .22\ X_1 + .14\ X_2$

$\sigma_I = 5.83$

(2.b) $\hat{G}_D = \frac{.22[\frac{1}{2}(80) + \frac{1}{4}(-40)] + .14[\frac{1}{4}(80) + 0\ ]}{5.83} D = 1.612\ D$

$\hat{G}_M = \frac{.22[\frac{1}{2}(-40) + \frac{1}{4}(40)] + .14[\frac{1}{4}(-40) + 0\ ]}{5.83} D = -.617\ D$

Selection when traits are influenced by grandmaternal as well as 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 large 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.

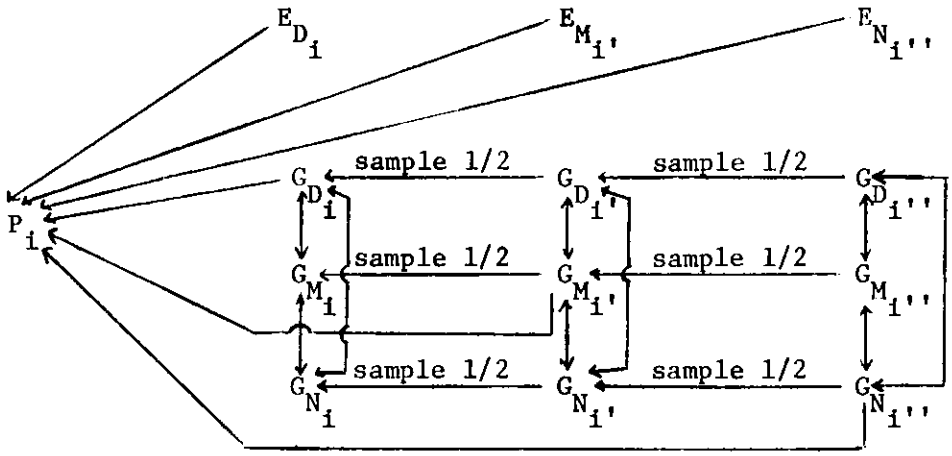
This grandmaternal effect may have a genetic basis in the grandmother (i'') but is an environmental effect on the maternal ability of the mother (i') and on the actual phenotype of the calf (i). In fact, the model including maternal effects can be expanded so that the maternal effect is made up of direct maternal effect and an environmental effect from the grandmother;

$$P_{M_{i'}} = G_{M_{i'}} + E_{i'} = G_{M_{i'}} + G_{N_{i''}} + E_{M_{i'}} + E_{N_{i''}},$$

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 some animal i can be expressed in increasingly partitioned form 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. A sample half of the genes are, of course, for the direct, maternal, and grandmaternal effects transmitted in each generation from parent to offspring.



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, we will define $G_X = G_{D_X} + G_{M_{X'}} + G_{N_{X''}}$ and $G_Y = G_{D_Y} + G_{M_{Y'}} + G_{N_{Y''}}$ although only 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 granddam of X, X''.

Thus,

$$\begin{aligned} E(G_X G_Y) = Cov(G_X G_Y) = {} & Cov(G_{D_X} G_{D_Y}) + Cov(G_{D_X} G_{M_{Y'}}) + Cov(G_{M_X}, G_{D_Y}) + \\ & Cov(G_{D_X} G_{N_{Y''}}) + Cov(G_{D_Y} G_{N_{X''}}) + Cov(G_{M_X}, G_{M_Y}) + \\ & Cov(G_{M_X}, G_{N_{Y''}}) + Cov(G_{M_Y}, G_{N_{X''}}) + Cov(G_{N_{X''}}, G_{N_{Y''}}). \end{aligned}$$

Each of these terms can be evaluated in terms of additive, dominance, additive by additive, etc., components of variance and covariance (where the direct, maternal, and grandmaternal components are considered separate traits).

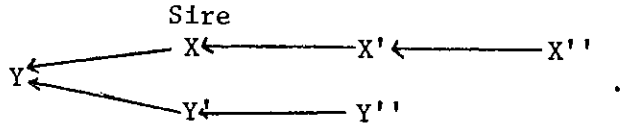
If only additive genetic effects are assumed, then

$$\begin{aligned} Cov(G_X G_Y) = {} & a_{XY}\sigma^2_{G_D} + (a_{XY'} + a_{YX'})\sigma_{G_D G_M} + (a_{XY''} + a_{YX''})\sigma_{G_D G_N} + \\ & a_{X'Y'}\sigma^2_{G_M} + (a_{X'Y''} + a_{Y'X''})\sigma_{G_M G_N} + a_{X''Y''}\sigma^2_{G_N}, \end{aligned}$$

where the variances are additive genetic variances and the covariances are

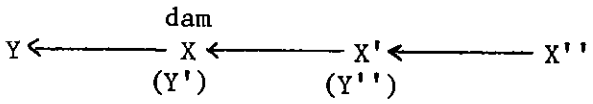
additive genetic covariances among the direct (D), maternal (M), and grandmaternal (N) effects.

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 grandams X'' and Y''. For example, if X is a sire and Y is the progeny, the diagram is



Thus, $a_{XY} = 1/2$, $a_{XY'} = 0$, $a_{YX'} = 1/4$, $a_{XY''} = 0$, $a_{YX''} = 1/8$, $a_{X'Y'} = 0$, $a_{X'Y''} = 0$, $a_{Y'X''} = 0$, and $a_{X''Y''} = 0$.

However, if X is a dam and Y is the progeny, the relationships are different.



Thus, $a_{XY} = 1/2$, $a_{XY'} = 1$, $a_{YX'} = 1/4$, $a_{XY''} = 1/2$, $a_{YX''} = 1/8$, $a_{X'Y'} = 1/2$, $a_{X'Y''} = 1$, $a_{Y'X''} = 1/4$, and $a_{X''Y''} = 1/2$. The relationships which are coefficients of the variances and covariances for some common relatives are given below.

	Component					
Relatives	$\sigma^2_{G_D}$	$\sigma_{G_D G_M}$	$\sigma_{G_D G_N}$	$\sigma^2_{G_M}$	$\sigma_{G_M G_N}$	$\sigma^2_{G_N}$
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	1/2	1	1	1
Sire, progeny	1/2	1/4	1/8	0	0	0
Dam, progeny	1/2	$1\frac{1}{4}$	5/8	1/2	$1\frac{1}{4}$	1/2
Full sibs	1/2	1	1/2	1	1	1
Maternal sibs	1/4	1	1/2	1	1	1
Paternal sibs	1/4	0	0	0	0	0
Granddam, grandprogeny	1/4	5/8	$1\frac{1}{16}$	1/4	5/8	1/4

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.

Selection may be for $T = G_{D\alpha}$ (additive direct for animal α) so that the right-hand sides become

$$\sigma_{X_i T} = E[(G_{D_{X_i}} + G_{M_{X_i'}} + G_{N_{X_i''}} + \text{other } E_{X_i})(G_{D\alpha})]$$
$$= a_{i\alpha}\sigma^2_{G_D} + a_{i'\alpha}\sigma_{G_D G_M} + a_{i''\alpha}\sigma_{G_D G_N}$$

and $\sigma^2_T = \sigma^2_{G_D}$. If $T = G_{M\alpha}$, $\sigma_{X_i T} = a_{i\alpha}\sigma_{G_D G_M} + a_{i'\alpha}\sigma^2_{G_M} + a_{i''\alpha}\sigma_{G_M G_N}$ and $\sigma^2_T = \sigma^2_{G_M}$. If $T = G_{N\alpha}$, $\sigma_{X_i T} = a_{i\alpha}\sigma_{G_D G_N} + a_{i'\alpha}\sigma_{G_M G_N} + a_{i''\alpha}\sigma^2_{G_N}$ 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

$$\hat{T} = v_D \hat{G}_{D\alpha} + v_M \hat{G}_{M\alpha} + v_N \hat{G}_{N\alpha} ,$$

or the selection index weights can be determined directly using the RHS's;

$$\sigma_{X_i T} = E[(G_{D_{X_i}} + G_{M_{X_i'}} + G_{N_{X_i''}} + \text{other E's})(v_D G_{D\alpha} + v_M G_{M\alpha} + v_N G_{N\alpha})]$$
$$= v_D(a_{i\alpha}\sigma^2_{G_D} + a_{i'\alpha}\sigma_{G_D G_M} + a_{i''\alpha}\sigma_{G_M G_N}) + v_M(a_{i\alpha}\sigma_{G_D G_M} + a_{i'\alpha}\sigma^2_{G_M} +$$
$$a_{i''\alpha}\sigma_{G_M G_N}) + v_N(a_{i\alpha}\sigma_{G_D G_N} + a_{i'\alpha}\sigma_{G_M G_N} + a_{i''\alpha}\sigma^2_{G_N})$$

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 $a_{i\alpha}$, $a_{i'\alpha}$, and $a_{i''\alpha}$ are not proportional) are necessary for selection with different economic values for the direct, maternal, and grandmaternal components.

FETAL EFFECTS MODEL (Sire of Fetus Effect)

There are some traits of a female which may be influenced by the fetus she is carrying either during the gestation or following the 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 can most certainly be partially influenced by the genes it carries. There is speculation that in dairy cattle, hormones secreted by the calf may influence the development of secretory tissue and thus influence milk production during the last part of gestation or during the lactation which follows birth of the fetus.

The 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. The figure shows that the animal contributes a sample 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.

1. Model

Figure 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 as described by Bar-Anan *et al.* (1976). In fact, any trait which is influenced by the mate of the female can be described by such a model. Fixed effects on the records will be ignored

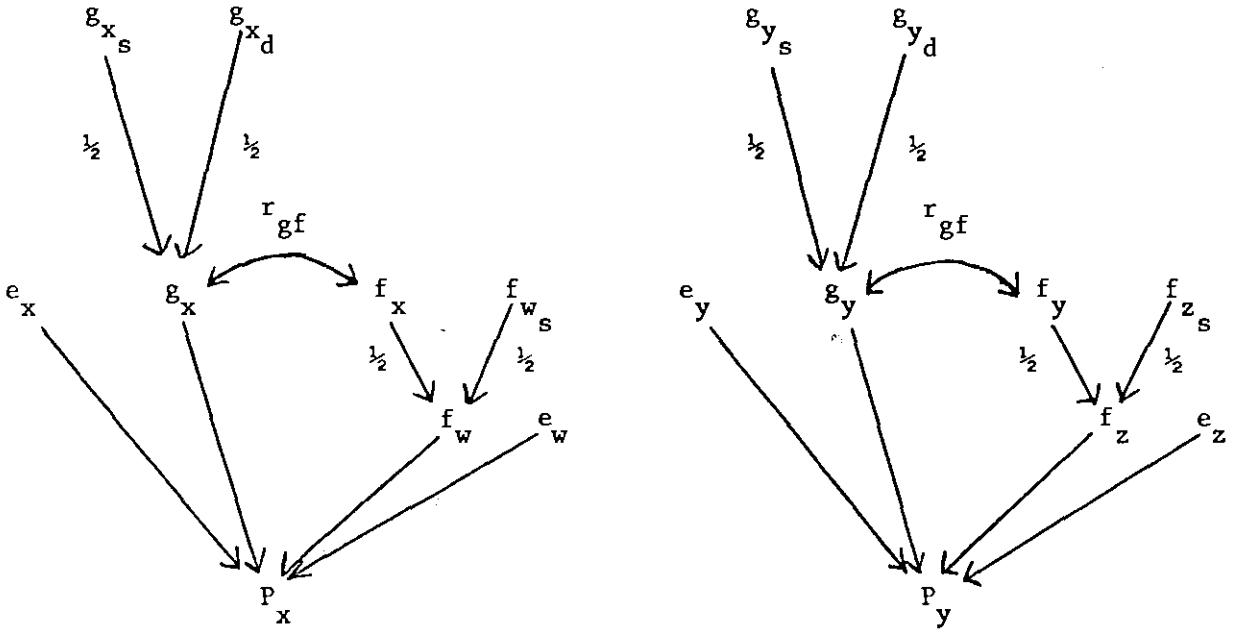


Figure 1. Diagram of direct genetic and environmental effects (g_x and e_x) and fetal genetic and environmental effects (f_w and e_w) on the phenotypic record of animal, x, carrying fetus, w. x_s, x_d are the sire and dam of x, w_s is the sire of the fetus, and of course, x is the dam of the fetus. A similar diagram is given for any potential relative, y, carrying fetus, z.

here but would need to be considered in prediction procedures or in estimation of components of variance.

Linear models including effects shown in Figure 1 are the same as used by Willham (1963). His application was to a maternal effects model.

The model is

$$P_x = g_x + f_w + e_x + e_w \quad \text{and} \quad P_y = g_y + f_z + e_y + e_z \qquad [1]$$

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 genetically determined.

$$\text{Cov}(P_x P_y) = \text{Cov}(g_x g_y) + \text{Cov}(f_w f_z) + \text{Cov}(g_x f_z) + \text{Cov}(g_y f_w)$$

and if only additive genetic effects are considered or assumed important, the covariances can be written as by Willham (1963):

$$\text{Cov}(P_x P_y) = a_{xy}\sigma_g^2 + a_{wz}\sigma_f^2 + (a_{xz} + a_{yw})\sigma_{gf} \qquad [2]$$

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.

This 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 a record of a dam and a record of her daughter when the dam's record was made with the influence of the fetus which was 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

$$\text{Cov(daughter-dam)} = \tfrac{1}{2}\sigma_g^2 + \tfrac{1}{2}\sigma_f^2 + (1 + \tfrac{1}{4})\sigma_{gf}$$

Expectations of covariances between usual combinations of records are given in Table 1.

Table 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[1] 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	1/2	1/8	1/2
Daughter-dam	Daughter from service sire of dam	1/2	1/2	5/4
Full sibs	Different	1/2	1/8	1/2
Full sibs	Same	1/2	3/8	1/2
Paternal or maternal sibs	Different	1/4	1/16	1/4
Paternal or maternal sibs	Same	1/4	5/16	1/4
Maternal sibs	Sire of x is service sire of y	1/4	3/16	1/2
Unrelated	Same	0	1/4	0

[1] Note that these covariances may also include other more likely components due to effects such as direct dominance and maternal additive.

The practical implications are that the effect of the sire of the cow includes the value of the sample half of his genes concerned directly with production and a sample quarter of his genes associated with the fetal effect since he is the grandsire of every calf his daughter produces. Thus,

Sire of cow effect (which we normally think of as the sire comparison)

= G/2 for production of cow + F/4 for production of cow.

The mate effect, or the fetal effect of the sample 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 cow)

= F/2 for production of cow

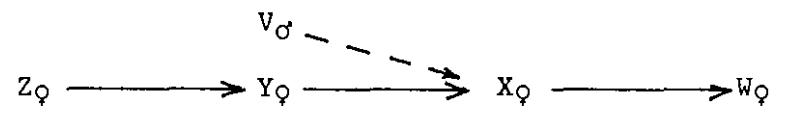
Note, however, that the sire of fetus is also the sire of the possible replacement heifer, resulting from birth and survival of the fetus. Thus, in the next generation, the sire of the fetus has become the sire of the cow. If there is a negative relationship between the direct and fetal effects then 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 the production of the mate, or for his direct genetic value, which does not become expressed until the resulting offspring becomes productive.

CYTOPLASMIC EFFECTS MODEL

The cytoplasm of the fertilized ovum comes primarily from the mother. Mitochondria in the cytoplasm are responsible for much of the cellular metabolism. The DNA of mitochondria in most species is inherited primarily from the mother. Thus, cytoplasmic effects generally are thought of as being maternal in origin and essentially to be unchanging along the maternal line. Males will express the cytoplasmic effects received from their mothers but will not transmit their cytoplasm to their offspring.

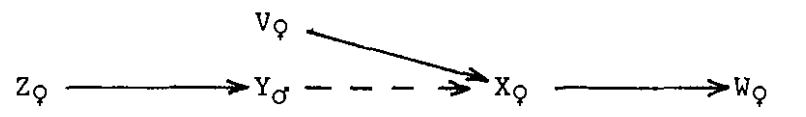
FIGURE 1. FRACTION OF ADDITIVE GENETIC (g) AND CYTOPLASMIC (c) EFFECTS IN DESCENDANTS.

For female line of descent:



g_z:	g_z	$g_z/2$	$g_z/4$	$g_z/8$
c_z:	c_z	c_z	c_z	c_z

With male in line of descent:



g_z:	g_z	$g_z/2$	$g_z/4$	$g_z/8$
c_z:	c_z	c_z	c_v	c_v

Cytoplasmic effects can be incorporated easily into selection index procedures either in computing the variances and covariances among the X's or the right-hand sides for selection of a function of direct additive 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 put in the model rather easily.

The models with cytoplasmic effects for records on relatives x and y are:

$$P_x = g_x + c_f + b_{x_f} + e_x \quad \text{and} \quad P_y = g_y + c_{f'} + b_{y_{f'}} + e_y$$

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 $f = f'$,

$$\text{cov}(c_f,c_{f'}) = \sigma_c^2, \text{ and } = 0 \text{ otherwise, and}$$

$$\text{cov}(b_{x_f},b_{y_{f'}}) = a_{xy}\sigma_b^2 \text{ and } 0 \text{ otherwise.}$$

As before, let $c_{xy}\sigma_x^2$ be the environmental covariance between a record of x and a record of 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 $f \neq f'$:

$$\text{cov}(P_x, P_y) = a_{xy}\sigma_g^2 + c_{xy}\sigma_x^2.$$

The following table gives the expected make-up of covariances of common relatives.

Relationship	σ^2_g	σ^2_c	σ^2_b	Environmental covariance $/\sigma^2_x$
Female parent - offspring	1/2	1	1/2	$c_{FP,O}$
Male parent - offspring	1/2	0	0	$c_{MP,O}$
Maternal half sibs	1/4	1	1/4	c_{MHS}
Paternal half sibs	1/4	0	0	c_{PHS}
Full sibs	1/2	1	1/2	c_{FS}
Female grandparent - offspring	1/4	1	1/4	$c_{FG,O}$
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^2_g, σ^2_c, σ^2_b, and c_{xy} are known, then variances, of, and covariances, among, averages can be calculated for setting up the coefficients of the selection index equations to find the selection index weights.

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:

$$\sigma_{X_i T} = a_{i\alpha}\sigma^2_g$$

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

if $f = f'$;

$$\sigma_{X_i T} = a_{i\alpha}\sigma^2_g + \sigma^2_c + a_{i\alpha}\sigma^2_b \text{ and}$$

if $f \neq f'$;

$$\sigma_{X_i T} = a_{i\alpha}\sigma^2_g$$

In general, for $T = g_\alpha + c_\alpha + b_{\alpha_{f'}}$;

$$\sigma_{X_i T} = a_{i\alpha}\sigma^2_g + P(f=f')[\sigma^2_c + a_{i\alpha}\sigma^2_b]$$

where P(f=f') is the probability that the cytoplasm of relative i with the record X_i and the animal being evaluated, α, is the same.

If $T = g_\alpha + c_\alpha$, then

if $f = f'$

$$\sigma_{x_i T} = a_{i\alpha}\sigma^2_g + \sigma^2_c \text{ and}$$

if $f \neq f'$

$$\sigma_{x_i T} = a_{i\alpha}\sigma^2_g$$

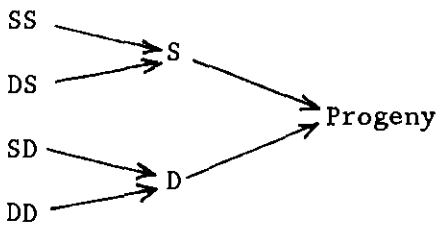
Unless σ^2_c is relatively large, selection for direct additive genetic value while ignoring cytoplasmic effects is likely to be nearly as effective as jointly selecting for direct additive and cytoplasmic effects.

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.

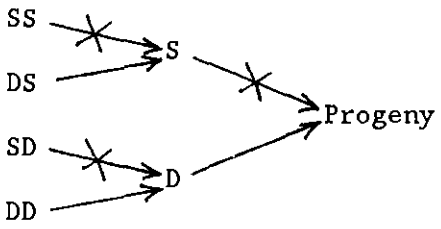
Unless variance due to cytoplasmic effects is large, the only way that selecting for cytoplasmic in addition to direct additive genetic value can be relatively important to total genetic gain is if the reproductive rate of females to produce female replacements is greatly increased. The reason is that 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 2. THE FOUR PATHS OF SELECTION

Nuclear inheritance



Cytoplasmic inheritance



For the sire to sire, dam to sire, and sire to dam paths, selection should be for additive genetic value with selection differentials of ΔSS, ΔDS, ΔSD for additive genetic value.

For the dam to dam path, selection can be for the 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 correlated response. If I is the index for the sum, g + c, then

$$\Delta DD_g = \frac{cov(g,I)}{\sigma_I} D$$

and

$$\Delta DD_c = \frac{cov(c,I)}{\sigma_I} D$$

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--daughter, dam, maternal granddam, maternal half sisters, full sibs, etc.

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, 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 cows 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 is likely to be greater than 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: COVARIANCE BETWEEN RELATIVES WITH SINGLE LOCUS FOR ADDITIVE EFFECTS AND MATERNALLY DERIVED CYTOPLASMIC EFFECT.

Let records of relatives x and y be represented as:

$$x_{ijt} = \alpha_i + \alpha_j + \gamma_t + (\alpha\gamma)_{it} + (\alpha\gamma)_{jt} + e_x$$

$$y_{k\ell u} = \alpha_k + \alpha_\ell + \gamma_u + (\alpha\gamma)_{ku} + (\alpha\gamma)_{\ell u} + e_y$$

where each α_m represents an additive genetic effect of gene m,

γ_n represents a cytoplasmic effect of cytoplasm n,

$(\alpha\gamma)_{mn}$ 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; $c_t = \gamma_t$ with $\sigma_c^2 = E[\gamma_t^2]$

genetic by cytoplasmic interaction; $b_{ijt} = (\alpha\gamma)_{it} + (\alpha\gamma)_{jt}$

with $\sigma_b^2 = E[(\alpha\gamma)_{it}^2] + E[(\alpha\gamma)_{jt}^2]$ 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 (either 1 or 0).

Cov(x,y):

$$E[g_{ij}g_{k\ell}] = E[(\alpha_i + \alpha_j)(\alpha_k + \alpha_\ell)] = E[\alpha_i\alpha_k + \alpha_i\alpha_\ell + \alpha_j\alpha_k + \alpha_j\alpha_\ell]$$

But $E[\alpha_i\alpha_k] = (a_{xy}/2)E[\alpha^2] = (a_{xy}/4)\sigma_g^2$

Thus, $E[g_{ij}g_{k\ell}] = a_{xy}\sigma_g^2$

$E[c_t c_u) = E[\gamma_t\gamma_u] = P(t=u)\sigma_c^2$; either σ_c^2 or 0.

$$E[b_{ijt}b_{k\ell u}] = E\{[(\alpha\gamma)_{it} + (\alpha\gamma)_{jt}][(\alpha\gamma)_{ku} + (\alpha\gamma)_{\ell u}]\}$$

$$= E[(\alpha\gamma)_{it}(\alpha\gamma)_{ku} + (\alpha\gamma)_{it}(\alpha\gamma)_{\ell u} + (\alpha\gamma)_{jt}(\alpha\gamma)_{ku} + (\alpha\gamma)_{jt}(\alpha\gamma)_{\ell u}]$$

But for t = u; $E[(\alpha\gamma)_i(\alpha\gamma)_k] = (a_{xy}/2)\ E[(\alpha\gamma)^2] = (a_{xy}/4)\sigma_b^2$

Thus, $E[b_{ijt}b_{k\ell u}] = P(t=u)a_{xy}\sigma_b^2$; either $a_{xy}\sigma_b^2$ or 0.

Therefore,

$$\mathrm{cov}(x,y) = a_{xy}\sigma^2_g + P(t=u)\sigma^2_c + P(t=u)a_{xy}\sigma^2_b$$

Reference

Beavis, W. D., E. Pollak, and K. J. Frey. 1987. A theoretical model for quantitatively inherited traits influenced by nuclear-cytoplasmic interactions. Theoretical and Applied Genetics 74:571-578.

Summary XIV

NONLINEAR ECONOMIC VALUES AND RESTRICTED SELECTION

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, i.e., the value of an additional pound of milk when the level is 109 lbs per day is not the same as when the level is 19 lbs. 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.

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.,

$$\text{Net income} = c + v_1(X_1+\mu_1) + v_2(X_1+\mu_1)^2 + v_3(X_1+\mu_1)^3 + \ldots ,$$

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 deviation from μ_1.

Thus, net genetic merit can be defined as

$$T = c + v_1(G_1+\mu_1) + v_2(G_1+\mu_1)^2 + v_3(G_1+\mu_1)^3 + \ldots ,$$

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 equations so that $\hat{T} = I = c + v_1(I_1+\mu_1) + v_2(I_1+\mu_1)^2 + v_3(I_1+\mu_1)^3 + \ldots$.

If only v_1 and v_2 are nonzero (linear and quadratic values), then this is the 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 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

$$\hat{T} = c' + I = c' + v_1(\mu_1+I_1) + v_2(\mu_2+I_2) + v_3(\mu_1+I_1)(\mu_2+I_2) + v_4(\mu_1+I_1)^2 + v_5(\mu_2+I_2)^2 ,$$

where c' is a constant for all $\hat{T}$. Wilton proved that this is equivalent to

$$I = c + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \beta_4 X_1^2 + \beta_5 X_2^2$$

where the β's are the solutions to the equations obtained from minimizing $E[(T-I) - E(T-I)]^2$.

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

$$I = \underbrace{[v_m + v_f(-\text{base test})]}_{v_1}(\mu_1+I_1) + \underbrace{v_f}_{v_3}(\mu_1+\ _1)(\mu_2+I_2) \ ,$$

where $I_1 = b_{11}X_1 + b_{12}X_2$ and $I_2 = b_{21}X_1 + b_{22}X_2$.

In some cases a simpler estimate 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 $E(X_1^2X_2)$.

An example of selection when milk price depends on fat test.

The example also demonstrates that an animal which ranks higher in one herd may not in another depending on the average milk yield and fat test.

Two sires have been evaluated for milk, I_m, and test, I_f. Two herds with widely different average milk and test are used.

Sire	I_m	I_f	Herd	μ_m	μ_f
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 following table gives the results for the six combinations of herds and pricing schemes.

	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

Suppose as an example with only two traits 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

$$T = a_0 + a_1T_1 + a_2T_2 + a_{12}T_1T_2 + a_{11}T_1 + a_{22}T_2 ,$$

where the constants $a_0 = v_1\mu_1 + v_2\mu_2 + v_{12}\mu_1\mu_2 + v_{11}\mu_1^2 + v_{22}\mu_2^2$, $a_1 = v_1 + v_{12}\mu_2 + 2v_{11}\mu_1$, $a_2 = v_2 + v_{12}\mu_1 + 2v_{22}\mu_2$, $a_{12} = v_{12}$, $a_{11} = v_{11}$, 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, $I = a_0 + a_1I_1 + a_2I_2 + a_{12}I_1I_2 + a_{11}I_1 + a_{22}I_2$, and $c = E(T) - E(I)$. c is the same constant 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;

$$\begin{aligned} E(T) &= E(a_0 + a_1T_1 + a_2T_2 + a_{12}T_1T_2 + a_{11}T_1 + a_{22}T_2) \\ &= a_0 + 0 + 0 + a_{12}\sigma_{T_1T_2} + a_{11}\sigma^2_{T_1} + a_{22}\sigma^2_{T_2} ; \\ E(I) &= E(a_0 + a_1I_1 + a_2I_2 + a_{12}I_1I_2 + a_{11}I_1 + a_{22}I_2) \\ &= a_0 + 0 + 0 + a_{12}E(I_1I_2) + a_{11}\sigma^2_{I_1} + a_{22}\sigma^2_{I_2} . \end{aligned}$$

$\sigma^2_{I_1} = r^2_{T_1I_1}\sigma^2_{T_1}$ 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,

$$E(T) - E(I) = a_{12}[\sigma_{T_1T_2} - E(I_1I_2)] + a_{11}\sigma^2_{T_1}(1-r^2_{T_1I_1}) + a_{22}\sigma^2_{T_2}(1-r^2_{T_2T_2}).$$

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 $E(I_1I_2) = b_{11}b_{21}\sigma^2_{X_1} + (b_{11}b_{22} + b_{12}b_{21})\sigma_{X_1X_2} + b_{12}b_{22}\sigma^2_{X_2}$.

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.

The general problem is to maximize $T = v_1G_1 + v_2G_2 + \ldots + v_MG_M$ but at the same time hold N - M other traits at their present genetic level, i.e., $\triangle_c G_{M+1} = 0 = \triangle_c G_{M+2} = \ldots = \triangle_c G_N$. A solution to this problem is given by Kempthorne and Nordskog in Biometrics (1959).

In the simplest case $T = v_1G_1$ and we want $\triangle_c G_2 = 0$. Available are measures on the two traits, X_1 and X_2. We will select for $T = v_1G_1$ 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 $\triangle T$ with the restriction that $\triangle_c G_2 = 0$.

The restriction, $\triangle_c G_2 = 0 = \frac{\mathrm{Cov}(G_2, I^*)}{\sigma_I^*} D$ so that $\mathrm{Cov}(G_2, I^*) = b_1^* \sigma_{X_1G_2} + b_2^* \sigma_{X_2G_2}$ must be zero.

In addition the equations for the b's to maximize r_{TI^*} are

$$\sigma_{X_1}^2 b_1^* + \sigma_{X_1X_2} b_2^* = \sigma_{X_1T}$$

$$\sigma_{X_1X_2} b_1^* + \sigma_{X_2}^2 b_2^* = \sigma_{X_2T}.$$

Thus there are three equations including the restriction but only two unknowns.

In order to find a solution we must add a dummy unknown--the so-called LaGrange multiplier, λ. The equations can now be solved and are symmetrical as before.

$$\begin{aligned} b_1^* \sigma_{X_1}^2 + b_2^* \sigma_{X_1X_2} + \lambda\sigma_{X_1G_2} &= \sigma_{X_1T} \\ b_1^* \sigma_{X_1X_2} + b_2^* \sigma_{X_2}^2 + \lambda\sigma_{X_2G_2} &= \sigma_{X_2T} \\ b_1^* \sigma_{X_1G_2} + b_2^* \sigma_{X_2G_2} + 0 &= 0 \end{aligned}$$

The restricted index will be $I^* = b_1^* X_1 + b_2^* X_2$. The λ can be found but is not needed.

These equations can be derived by minimizing $E[(T-I^*)^2]$ with the restriction that $2\lambda(b_1^*\sigma_{X_1G_2} + b_2^*\sigma_{X_2G_2}) = 0$, i.e., equate to zero the partial derivatives of $\sigma_T^2 + \sigma_{I^*}^2 - 2\sigma_{TI^*} + 2\lambda\sigma_{G_2I^*}$ 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 λ there will be λ_i, $i = M+1, \ldots, N$ where $N - M$ is the number of traits to hold constant and N is the number of economic traits.

As an example consider $M = 2, (T = v_1G_1 + v_2 G_2)$, and $N-M = 2$ ($\triangle_c G_3 = 0 = \triangle_c G_4$).

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 will be

$$\sigma_{G_3I^*} = 0 = \beta_1^* \sigma_{X_1G_3} + \beta_2^* \sigma_{X_2G_3} + \beta_3^* \sigma_{X_3G_3} + \beta_4^* \sigma_{X_4G_3} \quad \text{and}$$
$$\sigma_{G_4I^*} = 0 = \beta_1^* \sigma_{X_1G_4} + \beta_2^* \sigma_{X_2G_4} + \beta_3^* \sigma_{X_3G_4} + \beta_4^* \sigma_{X_4G_4} .$$

Thus, λ_1 and λ_2 will be the LaGrange multipliers in the equations to find the restricted selection index weights:

$$\beta_1^* \sigma_{X_1}^2 + \beta_2^* \sigma_{X_1X_2} + \beta_3^* \sigma_{X_1X_3} + \beta_4^* \sigma_{X_1X_4} + \lambda_1 \sigma_{X_1G_3} + \lambda_2 \sigma_{X_1G_4} = \sigma_{X_1T}$$
$$\beta_1^* \sigma_{X_1X_2} + \beta_2^* \sigma_{X_2}^2 + \beta_3^* \sigma_{X_2X_3} + \beta_4^* \sigma_{X_2X_4} + \lambda_1 \sigma_{X_2G_3} + \lambda_2 \sigma_{X_2G_4} = \sigma_{X_2T}$$
$$\beta_1^* \sigma_{X_1X_3} + \beta_2^* \sigma_{X_2X_3} + \beta_3^* \sigma_{X_3}^2 + \beta_4^* \sigma_{X_3X_4} + \lambda_1 \sigma_{X_3G_3} + \lambda_2 \sigma_{X_3G_4} = \sigma_{X_3T}$$
$$\beta_1^* \sigma_{X_1X_4} + \beta_2^* \sigma_{X_2X_4} + \beta_3^* \sigma_{X_3X_4} + \beta_4^* \sigma_{X_4}^2 + \lambda_1 \sigma_{X_4G_3} + \lambda_2 \sigma_{X_4G_4} = \sigma_{X_4T}$$
$$\beta_1^* \sigma_{X_1G_3} + \beta_2^* \sigma_{X_2G_3} + \beta_3^* \sigma_{X_3G_3} + \beta_4^* \sigma_{X_4G_3} + 0 + 0 = 0$$
$$\beta_1^* \sigma_{X_1G_4} + \beta_2^* \sigma_{X_2G_4} + \beta_3^* \sigma_{X_3G_4} + \beta_4^* \sigma_{X_4G_4} + 0 + 0 = 0$$

The response in selecting according to I* should probably be compared to the response in selecting directly for $T = v_1G_1 + v_2G_2$ by $I = \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4$ 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. It may be that although G_3 and G_4 are optimum that the restriction to maintain that optimum will be so costly in terms of $\triangle G_1$ and $\triangle G_2$ 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 this problem.

In the typical example above total response $\circledast T = \frac{Cov(T, I^*)}{\sigma_{I^*}} D = v_1 \circledast G_1 + v_2 \circledast G_2$ where as usual $\circledast G_1 = \frac{Cov(G_1, I^*)}{\sigma_{I^*}} D$ and $\circledast G_2 = \frac{Cov(G_2, I^*)}{\sigma_{I^*}} D$. $\circledast G_3 = \circledast G_4 = 0$. These would be compared with $v_1\triangle G_1 + v_2\triangle G_2 + v_3\triangle G_3 + v_4\triangle G_4$ as calculated from using the unrestricted index where v_3 and v_4 will depend on the loss in changing G_3 and G_4 from their present optimums. It may be that the values of traits 3 and 4 will be different when the changes are negative from when the changes are positive.

Examples using the restricted selection index

Records on animal being evaluated

Suppose milk yield, trait 1, is to be improved and fat test, trait 2, is to be held constant. A record on each trait is available on all animals to be evaluated.

Given: $\sigma^2_{P_1} = (2500 \text{ lb})^2$ $\sigma^2_{P_2} = (.3\%)^2$ $\sigma_{P_1P_2} = -150$

$\sigma^2_{G_1} = (1250 \text{ lb})^2$ $\sigma^2_{G_2} = (.21\%)^2$ $\sigma_{G_1G_2} = -157.5$

$r_{P_{12}} = -.2$, $r_{g_{12}} = -.6$, $h_1^2 = .25$, $h_2^2 = .49$.

Let $v_1 = 1$ since selection is for only one trait.

Find $I_1^* = b_1^* X_1 + b_2^* X_2$.

The equations determining the weights are:

$$(2500)^2\, b_1^* - 150\, b_2^* - 157.5\lambda = (1250)^2$$

$$-150\, b_1^* + (.3)^2\, b_2^* + (.21)^2\lambda = -157.5$$

$$-157.5\, b_1^* + (.21)^2\, b_2^* = 0 \quad \text{and}$$

$$I_1^* = .159\, X_1 + 566.67\, X_2$$

$$\sigma^2_{I_1^*} = (.159)^2\,(2500)^2 + (566.67)^2\,(.3)^2 + 2(.159)(566.67)(-150) = 159{,}876;$$

$$\sigma_{I^*} = 399.84.$$

$$\triangle^* G_1 = \frac{\mathrm{Cov}(G_1, I_1^*)}{\sigma_{I_1^*}} D = \frac{.159(1250)^2 + 566.67(-157.5)}{399.84} D = 398\, D \text{ (lb)}.$$

$$\triangle^* G_2 = \frac{\mathrm{Cov}(G_2, I_1^*)}{\sigma_{I_1^*}} D = \frac{.159(-157.50) + 566.67(.21)^2}{399.84} D = 0.$$

If selection is for G_1 with no restriction on G_2:

$$I_1 = .217\, X_1 - 1388.9\, X_2, \quad \sigma_{I_1} = 747,$$

$$\triangle G_1 = 747\, D \text{ (lb) and}$$

$$\triangle G_2 = \frac{\mathrm{Cov}(G_2, I_1)}{\sigma_{I_1}} D = \frac{.217(-157.5) - 1388.9(.21)^2}{747} D = -.1277\, D \text{ (\%)}.$$

If selection is for G_1 from X_1 only:

$$I_1 = .25X_1, \quad \sigma_I = 625,$$

$$\triangle G_1 = 625\, D \text{ (lb), and}$$

$$\triangle G_2 = \frac{.25(-157.5)}{625} D = -.063\, D \text{ (\%)}.$$

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 daughter average for milk, X_{12} is the daughter average for test with 1 record per daughter and p = 20 daughters.

I* (1) is to improve G_1 and not change G_2, I (2) to maximize $\triangle G_1$, I (3) to maximize $\triangle G_1$.

Given: $\sigma_{P_1} = 2500$ lb., $\sigma_{P_2} = .3\%$, $h_1^2 = .25$, $h_2^2 = .49$, $r_{g_{12}} = -.6$, $r_{P_{12}} = -.2$ as before.

Find the indexes and expected response in milk for all 3 procedures.

(1) $I^* = b_1^* X_{11} + b_2^* X_{12}$ thus want $\text{Cov}(G_{2\alpha}, I^*) = b_1^* a_{1\alpha} \sigma_{G_2G_1} + b_2^* a_{1\alpha} \sigma^2_{G_2} = 0$

Equations:

$$\left(\frac{1 + (p-1)\frac{1}{4}h_1^2}{p}\right)\sigma^2_{P_1} b_1^* + \frac{\sigma_{P_1P_2} + (p-1)\, a_{11'}\, \sigma_{G_1G_2}}{p} b_2^* + a_{1\alpha}\, \sigma_{G_1G_2}\lambda = a_{1\alpha}\, \sigma^2_{G_1}$$

$$\frac{\sigma_{P_1P_2} + (p-1)\, a_{11'}\, \sigma_{G_1G_2}}{p}\, b_1^* + \left(\frac{1 + (p-1)\frac{1}{4}h_2^2}{p}\right)\sigma^2_{P_2} b_2^* + a_{1\alpha}\, \sigma^2_{G_2}\lambda = a_{1\alpha}\, \sigma_{G_1G_2}$$

$$a_{1\alpha}\, \sigma_{G_1G_2}\, b_1^* + a_{1\alpha}\, \sigma^2_{G_2}\, b_2^* + 0\,\lambda = 0$$

Numerically:

$$683{,}594\, b_1^* - 45\, b_2^* - 78.75\,\lambda = 781{,}250$$

$$-45\, b_1^* + .01497375\, b_2^* + .02205\lambda = -78.75$$

$$-78.75\, b_1^* + .02205\, b_2^* = 0$$

$$I^* = .9047\, X_{11} + 3231\, X_{12}$$

$$\sigma^2_{I^*} = 452{,}748$$

$$\sigma_{I^*} = 673$$

$$\text{⚠}^* G_1 = \frac{\text{Cov}(G_{\alpha 1}, I^*)}{\sigma_{I^*}} D = \frac{.9047(\frac{1}{2}\sigma^2_{G_1}) + 3231(\frac{1}{2}\sigma_{G_1G_2})}{673} D = 672\, D \text{ (lb)}$$

$$\text{⚠}^* G_2 = \frac{\text{Cov}(G_{\alpha 2}, I^*)}{\sigma_{I^*}} D = \frac{.9047(\frac{1}{2}\sigma_{G_1G_2}) + 3231(\frac{1}{2}\sigma^2_{G_2})}{673} D = 0\ (\%)$$

(2) $I = b_1 X_{11} + b_2 X_{12}$: equations are upper 2×2 for (1) and same 2 RHS's.

$I = .9943\ X_{11} - 2258\ X_{12} \quad \sigma_I^2 = 954{,}229, \quad \sigma_I = 977$

$$\triangle_c G_1 = \frac{\mathrm{Cov}(G_{\alpha 1},\ I)}{\sigma_I} D = \frac{.9943(\frac{1}{2}\sigma_{G_1}^2) - 2258(\frac{1}{2}\sigma_{G_1G_2})}{977} D = 977\ D \text{ (lb)}$$

$$\triangle_c G_2 = \frac{\mathrm{Cov}(G_{\alpha 2},\ I)}{\sigma_I} D = \frac{.9943(\frac{1}{2}\sigma_{G_1G_2}) - 2258(\frac{1}{2}\sigma_{G_2}^2)}{977} D = -.1311\ D\ (\%)$$

(3) $I = b_1 X_{11}$: equation is first diagonal and first RHS of (1) or (2).

$I = 1.1429\ X_{11} \quad \sigma_I^2 = 892{,}924 \quad \sigma_I = 945$

$$\triangle_1 G_1 = \frac{\mathrm{Cov}(G_{\alpha 1},\ I)}{\sigma_I} D = \frac{1.1429(\frac{1}{2}\sigma_{G_1}^2)}{945} D = 945\ D \text{ (lb)}$$

$$\triangle_1 G_2 = \frac{\mathrm{Cov}(G_{\alpha 2},\ I)}{\sigma_I} D = \frac{1.1429(\frac{1}{2}\sigma_{G_1G_2})}{945} D = -.0952\ D\ (\%)$$

Procedure	Cow selection		Bull selection	
	$\triangle G_1/D$	$\triangle G_2/D$	$\triangle G_1/D$	$\triangle G_2/D$
$I_1^* = b_1^* X_{11} + b_2^* X_{12}$	398 lb	0 %	672 lb	0 %
$I_1 = b_1 X_{11} + b_2 X_{12}$	747 lb	-.13 %	977 lb	-.13 %
$I_1 = b_1 X_{11}$	625 lb	-.06 %	945 lb	-.10 %

Summary XV

INDEX AND ECONOMIC VALUES IN RETROSPECT AND SELECTION EMPHASIS

The index in retrospect

The index in retrospect is an index that has been used for selection even though the weights were unknown at the time of selection. Determining the index that was used depends on finding the index which would give a particular set of phenotypic selection differentials.

Let $I = \sum_{i=1}^{N} w_i P_i$ be the underlying but unknown index that was used for selection and D the selection intensity factor. P_i will be the phenotypic record for trait i measured on the animal being selected.

With respect to the underlying unknown index, I, the regression of P_j on I gives the expected phenotypic selection differential for trait j:

$$\Delta P_j = \frac{\mathrm{Cov}(P_j,I)}{\sigma_I^2} \Delta I = \frac{\mathrm{Cov}(P_j,I)}{\sigma_I} D = (D/\sigma_I)\ \mathrm{Cov}(P_j,I) \ .$$

D/σ_I will be a constant for all traits, thus not changing the proportionality of the right-hand sides for different traits although both D and I may be different for males and females and even from generation to generation. Indexes in retrospect may be computed separately for males and females for each generation. If D/σ_I is set to one, then the expectations of the phenotypic selection differentials are:

$$\begin{aligned}
\Delta P_1 &= \mathrm{Cov}(P_1,I) = w_1\sigma_{P_1}^2 + w_2\sigma_{P_1P_2} + \ldots + w_N\sigma_{P_1P_N} \\
\Delta P_2 &= \mathrm{Cov}(P_2,I) = w_1\sigma_{P_1P_2} + w_2\sigma_{P_2}^2 + \ldots + w_N\sigma_{P_2P_N} \\
&\ \ \vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \\
\Delta P_N &= \mathrm{Cov}(P_N,I) = w_1\sigma_{P_1P_N} + w_2\sigma_{P_2P_N} + \ldots + w_N\sigma_{P_N}^2 \ .
\end{aligned}$$

Note that the coefficients of the w's are the same as for finding the best index weights, i.e., the phenotypic variances and covariances. Equating the

selection differentials to these equations will determine in retrospect the relative weights used in the index. The phenotypic variances and covariances must be known.

The proportionality of the w's does not appear to depend on the selection intensity or σ_I. A linear index in the phenotypic values is assumed as is truncation selection based on the underlying but unknown index.

The expected correlated responses from using the retrospective index are $\Delta G_j = (D/\sigma_I)\ \mathrm{Cov}(G_j,I)$ [j=1, ..., N] which can be compared with the expected responses from the theoretically best index.

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 would be estimated as the difference in phenotypic means between 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:

$$\Delta G_j = \frac{\mathrm{Cov}(G_j,I)}{\sigma_I^2}\,\Delta I = \frac{\mathrm{Cov}(G_j,I)}{\sigma_I}\,D = (D/\sigma_I)\ \mathrm{Cov}(G_j,I)\ .$$

Set $D/\sigma_I = 1$, and the expected values of the genetic selection differentials are:

$$\begin{aligned}
\Delta G_1 &= \mathrm{Cov}(G_1,I) = w_1\sigma^2_{G_1} + w_2\sigma_{G_1G_2} + \ldots + w_N\sigma_{G_1G_N}\\
\Delta G_2 &= \mathrm{Cov}(G_2,I) = w_1\sigma_{G_1G_2} + w_2\sigma^2_{G_2} + \ldots + w_N\sigma_{G_2G_N}\\
&\ \ \vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots\\
\Delta G_N &= \mathrm{Cov}(G_N,I) = w_1\sigma_{G_1G_N} + w_2\sigma_{G_2G_N} + \ldots + w_N\sigma^2_{G_N}\ .
\end{aligned}$$

Thus, if the genetic variance-covariance matrix is known as well as the genetic selection differentials, the weights for the underlying index are the solutions to the above equations.

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 $T = \Sigma v_i G_i$. Thus, the usual equations to find the weights (which are now known) can be used to find the corresponding economic values. The calculated numerical values on the left-hand sides are equated to the right-hand sides.

$$\begin{aligned}
\sigma^2_{P_1} w_1 + \sigma_{P_1P_2} w_2 + \ldots + \sigma_{P_1P_N} w_N &= v_1\sigma^2_{G_1} + v_2\sigma_{G_1G_2} + \ldots + v_M\sigma_{G_1G_N} \\
\sigma_{P_1P_2} w_1 + \sigma^2_{P_2} w_2 + \ldots + \sigma_{P_2P_N} w_N &= v_1\sigma_{G_2G_1} + v_2\sigma^2_{G_2} + \ldots + v_M\sigma_{G_2G_N} \\
\vdots \qquad\quad \vdots \qquad\qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\qquad \vdots & \\
\sigma_{P_1P_N} w_1 + \sigma_{P_2P_N} w_2 + \ldots + \sigma^2_{P_N} w_N &= v_1\sigma_{G_NG_1} + v_2\sigma_{G_NG_2} + \ldots + v_M\sigma^2_{G_N}
\end{aligned}$$

Even if the LHS's are unknown, this would be equivalent to equating the phenotypic selection differentials to the right-hand sides of the selection index equations for predicting total merit and then solving for the economic values. The genetic variances and covariances are necessary for determining the economic values.

$$\begin{aligned}
\Delta P_1 &= v_1\sigma^2_{G_1} + \ldots + v_N\sigma_{G_1G_N} \\
\vdots \qquad &\quad \vdots \qquad\qquad \vdots \\
\Delta P_N &= v_1\sigma_{G_NG_1} + \ldots + v_N\sigma^2_{G_N} .
\end{aligned}$$

An empirical selection index

If the net value of each animal can be determined (even with error uncorrelated with the X's), then the empirical selection index, $I = \Sigma\beta_i X_i$, can be found from the multiple regression of net value, y, on the phenotypic traits,

X_i (i=1, ..., N):

$$\sigma^2_{X_1}\beta_1 + \sigma_{X_1X_2}\beta_2 + \ldots + \sigma_{X_1X_N}\beta_N = \sigma_{X_1y}$$
$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$
$$\sigma_{X_NX_1}\beta_1 + \sigma_{X_NX_2}\beta_2 + \ldots + \sigma^2_{X_N}\beta_N = \sigma_{X_Ny}$$

The variances and covariances are all estimated from the data which includes net value. The phenotypic variances and covariances could be estimated from a larger sample of data, some of which does not include net value. Solving these equations will give the empirical selection index which is an unbiased estimate of the best index to predict overall economic value.

Economic values from empirical covariances

The model for net value for animal m is $y_m = \Sigma v_j G_{jm} + e_m$ 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 (not a very reasonable assumption), then $\sigma_{X_iy} = \sum_j v_j \sigma_{G_iG_j}$ for all i which would give these equations:

$$v_1\sigma^2_{G_1} + v_2\sigma_{G_1G_2} + \ldots + v_N\sigma_{G_1G_N} = \sigma_{X_1y}$$
$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$
$$v_1\sigma_{G_NG_1} + v_2\sigma_{G_NG_2} + \ldots + v_N\sigma^2_{G_N} = \sigma_{X_Ny} .$$

If the genetic variances and covariances are known and the σ_{X_1y} have been computed, the equations can be solved to find the economic values.

If the empirical selection index weights are unbiased and since the right-hand sides should be $\Sigma v_j \sigma_{G_iG_j}$ for all i, the economic values can also be estimated from these equations since $I = \Sigma\beta_i X_i$ is a retrospective index:

$$\sigma^2_{X_1}\beta_1 + \ldots + \sigma_{X_1X_N}\beta_N = v_1\sigma^2_{G_1} + \ldots + v_N\sigma_{G_1G_N}$$
$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$
$$\sigma_{X_NX_1}\beta_1 + \ldots + \sigma^2_{X_N}\beta_N = v_1\sigma_{G_NG_1} + \ldots + v_N\sigma^2_{G_N}$$

If only the v's are unknown, the equations can be solved to find the economic values.

Example of calculation of relative selection emphasis on traits other than a major trait in selected matings

Cows for selected matings are often chosen from a list of the highest 5% of cows evaluated for milk production when the highest 1% would be enough to produce the required number of young bulls. Obviously traits other than milk production are being considered. Some idea of the relative selection emphasis on milk and other traits would be informative to those involved in sire selection.

Expected progress from selection for one trait can be written as $\Delta G = r_{TI} D \sigma_G$ where r_{TI} is the accuracy of evaluation, D is the selection intensity factor, and σ_G is the genetic standard deviation. The genetic standard deviation is a constant for a trait. If r_{TI} is assumed to be the same, then the relative genetic progress for two selection intensities can be written as D_2/D_1 where D_1 is the selection intensity factor for the highest selection intensity and D_2 is the other selection intensity factor.

If all traits are assumed to have equal heritabilities, h^2, are uncorrelated, and are standardized, then expected progress for a trait such as milk can be calculated easily when selection is for milk and other traits.

Suppose that selection is for milk, y_1, and another trait, y_2, with relative selection emphasis of 1:m for milk and the other trait. Then the index is proportional to $I_m = y_1 + my_2$. The progress for milk when selecting for milk from $I = y_1$ can be written as $h^2 D_1$. Progress for milk when selecting by $I_m = y_1 + my_2$ can be written as

$$\frac{\mathrm{Cov}(I_m, G_1)}{\sigma_{I_m}} D_1 = h^2 D_1 / \sqrt{1 + m^2}$$

which results in relative progress for milk of

$$\frac{h^2 D_1 / \sqrt{1 + m^2}}{h^2 D_1} = \frac{1}{\sqrt{1 + m^2}} .$$

This relative progress can be equated to the same relative progress that would result from decreased selection intensity, i.e., a selection intensity factor of D_2 rather than D_1;

$$\frac{1}{\sqrt{1 + m^2}} = \frac{D_2}{D_1} .$$

From this, $m = \sqrt{(D_1/D_2)^2 - 1}$, where m is the relative selection emphasis on the trait other than milk.

Similarly, if selection is for milk, y_1, and n other traits, $y_2 \ldots y_{n+1}$, with equal selection emphasis on all traits, then the index is proportional to $I_n = y_1 + y_2 + \ldots + y_{n+1}$. Expected progress for milk when selecting for the n+1 traits can be written as

$$\frac{\mathrm{Cov}(I_n, G_1)}{\sigma_{I_n}} D_1 = h^2 D_1/\sqrt{1 + n}$$

with relative progress for milk of

$$\frac{h^2 D_1/\sqrt{1 + n}}{h^2 D_1} = \frac{1}{\sqrt{1 + n}} .$$

Again this can be equated to relative progress for decreased selection intensity;

$$\frac{1}{\sqrt{1 + n}} = \frac{D_2}{D_1}$$

which yields $n = (D_1/D_2)^2 - 1$, where n is the number of traits with the same selection emphasis as for milk yield.

The table on the next page gives the m and n values corresponding to decreasing selection intensity when enough matings can be made from the top 1% of cows evaluated for milk yield.

Table . m and n values corresponding to decreasing selection intensity when enough matings can be made from the top 1% of cows evaluated for milk yield.

Cows selected from this fraction when only .01 needed	Relative emphasis (1:m) on milk and one other trait	Number (n) of other traits with same emphasis
.01	1 : 0.000	0.000
.05	1 : 0.538	0.289
.10	1 : 0.718	0.516
.20	1 : 0.949	0.900
.30	1 : 1.138	1.295
.40	1 : 1.324	1.754
.50	1 : 1.527	2.333
.60	1 : 1.769	3.130
.70	1 : 2.086	4.352
.80	1 : 2.569	6.600
.90	1 : 3.530	12.641

SUMMARY XVI

PREDICTION FROM LINEAR MIXED MODELS

Selection index procedures described in previous sections 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 represents 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 use 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.

The mixed model procedure was derived by C. R. Henderson about 1948. He has generalized and proved its properties since that time. The procedure results in what is called best linear unbiased prediction (BLUP), where 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 mixed model equations.

Both BLUP and selection index procedures require the assumption that variances such as genetic and phenotypic variances are known. The properties in common 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 pair-wise 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 errors of estimates of fixed effects with the condition 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.

When all observations have the same variance, the procedure simplifies to a simple set of equations involving all effects in the model except for the residual effects. (The procedure is considerably more complex with multiple traits with different variances and covariances. Multiple trait applications

will not be discussed.) The number of equations is the same as the number of effects in the model.

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.

The 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).

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. Write down the model for each sum (that is, the expected value of the sum considering all effects as fixed) excluding the residual term. Equate each sum to its model. 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_v^2, then add the ratio, σ_w^2/σ_v^2, to the diagonal coefficient of those equations. σ_w^2 is the variance of residual terms. 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 fixed effects equations. The rule is one nonestimable constraint 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 effects in all classifications except one also set equal to 0 if there are more than one fixed classification.

Interpretation of solutions

1. Solutions for the 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) $E\begin{bmatrix}\text{difference in solutions for two fixed} \\ \text{effects in the same classification}\end{bmatrix} = \begin{matrix}\text{actual difference in} \\ \text{the fixed effects}\end{matrix}$

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, the dot notation will be introduced which makes writing the equations in a symbolic form less laborious. 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 $n_3 = 4$. The total number of records is $\sum_{i=1}^{3} n_i = n_1 + n_2 + n_3$. In dot notation this is written as $n_.$. Similarly, the sum of all records of animal 1 is

$$\sum_{j=1}^{n_1=2} P_{1j} = P_{11} + P_{12} = P_{1.} \quad .$$

The sum of all records is

$$\sum_{i=1}^{3} \sum_{j=1}^{n_i} P_{ij} = P_{11} + P_{12} + P_{21} + P_{31} + P_{32} + P_{33} + P_{34}$$
$$= P_{1.} + P_{2.} + P_{3.} = P_{..}$$

EXAMPLES

Example 1: One-way fixed classification model

Suppose records are classified by the age when the record is made and that each animal has only one record. Then the 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 include 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) starts at 1 for each age group (i).

Further suppose the following records are available (the record will be equated to its model to clarify rules 1 and 2).

$115 = y_{11} = \mu + A_1 + w_{11}$ $\qquad$ $95 = y_{21} = \mu + A_2 + w_{21}$

$85 = y_{12} = \mu + A_1 + w_{12}$ $\qquad$ $90 = y_{31} = \mu + A_3 + w_{31}$

$105 = y_{13} = \mu + A_1 + w_{13}$ $\qquad$ $110 = y_{32} = \mu + A_3 + w_{32}$

Thus, i = 1, 2, or 3; and $n_1 = 3$, the number of records for age group 1; $n_2 = 1$; and $n_3 = 2$.

Rule 1 states that a sum is to be computed for each effect in the model (excluding w terms). The four effects in the model are μ, A_1, A_2, and A_3. The sum for μ includes each record having μ in its model which is true for all records; Thus, $y_{..} = 600$. The sum for A_1 includes each record containing A_1 which is true for the n_1 records with subscript i = 1; thus, $y_{1.} = 115 + 85 + 105 = 305$. Similarly the sum for A_2 is $y_{2.} = 95$ and for A_3 is $y_{3.} = 90 + 110 = 200$.

The next step is to equate each sum to its model (excluding the w terms).

The model for $y_{..}$ is simply the sum of the models for all records, $n_.$, 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 the sum: all n_1 records contain μ and A_1 so that the model for $y_{1.}$ is $n_1\mu + n_1A_1$.

The same pattern applies to the model for $y_{2.}$ which is $n_2\mu + n_2A_2$ and for $y_{3.}$ which is $n_3\mu + n_3A_3$.

Written in their usual symbolic form and with ^'s to indicate solutions:

$$\begin{aligned}
\mu: &\quad n_{..}\hat{\mu} + n_1\hat{A}_1 + n_2\hat{A}_2 + n_3\hat{A}_3 = y_{..} \\
A_1: &\quad n_1\hat{\mu} + n_1\hat{A}_1 = y_{1.} \\
A_2: &\quad n_2\hat{\mu} + n_2\hat{A}_2 = y_{2.} \\
A_3: &\quad n_3\hat{\mu} + n_3\hat{A}_3 = y_{3.}
\end{aligned}$$

For the example the numerical equations 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., coefficients in the first row are the same as in the first column, etc.

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.

3) The three A equations sum to the μ equation. Thus, even though there are four equations in four unknowns, the equations are not independent. To obtain a set of solutions, one constraint must be imposed on the original four solutions. (See rule 4.)

a) The constraint $\hat{\mu} = 0$ is the easiest to use computationally. The equation for μ is eliminated (to maintain symmetry) as well as the $\hat{\mu}$ 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 constraint is to set $\hat{A}_3 = 0$; the equation for A_3 is eliminated (to maintain symmetry) as well as $\hat{A}_3$ in the remaining equations:

$$
\begin{aligned}
n_.\hat{\mu} + n_1\hat{A}_1 + n_2\hat{A}_2 &= y_{..} \\
n_1\hat{\mu} + n_1\hat{A}_1 &= y_{1.} \\
n_2\hat{\mu} + n_2\hat{A}_2 &= y_{2.}
\end{aligned}
$$

c) A more complex restraint 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 (LaGrange multiplier, λ) is added to each equation so that:

$$
\begin{aligned}
n_.\hat{\mu} + n_1\hat{A}_1 + n_2\hat{A}_2 + n_3\hat{A}_3 + 0\lambda &= y_{..} \\
n_1\hat{\mu} + n_1\hat{A}_1 + 1\lambda &= y_{1.} \\
n_2\hat{\mu} + n_2\hat{A}_2 + 1\lambda &= y_{2.} \\
n_3\hat{\mu} + n_3\hat{A}_3 + 1\lambda &= y_{3.} \\
0\hat{\mu} + 1\hat{A}_1 + 1\hat{A}_2 + 1\hat{A}_3 &= 0 \quad .
\end{aligned}
$$

The solutions with the constraint $\hat{\mu} = 0$ are the easiest to discuss:

$$\hat{\mu} = 0 \qquad \hat{A}_2 = y_{2.}/n_2$$
$$\hat{A}_1 = y_{1.}/n_1 \qquad \hat{A}_3 = y_{3.}/n_3$$

Note that the constraint $\hat{\mu} = 0$ is one of the solutions. Another result of having to impose a constraint is the property stated under 2 in the interpretation of solutions.

Obviously, and in many cases, $E[\hat{\mu}] \neq \mu$ since $E[0] = 0$. $E[\hat{A}_1]$ can be found easily with the one-way classification model. Note that $E[y_{ij}] = \mu + A_i$ for all j. Thus,

$$E[\hat{A}_1] = E[y_{1.}/n_1] = (1/n_1)\, E[y_{11} + y_{12} + y_{13} + \ldots + y_{1n_1}]$$
$$= (1/n_1)\, [\, n_1(\mu + A_1)]$$
$$= \mu + A_1 \quad .$$

Similarly

$$E[\hat{A}_2] = \mu + A_2 \qquad \text{and}$$
$$E[\hat{A}_3] = \mu + A_3 \quad .$$

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 $\mu + A_i$. For example, $A_1 - A_2$ can be estimated by $\hat{A}_1 - \hat{A}_2$ since $E[\hat{A}_1 - \hat{A}_2] = [(\mu + A_1) - (\mu + A_2)] = A_1 - A_2$. Solutions obtained using other constraints will have different expectations. For the constraint $\hat{A}_3 = 0$, $E[\hat{A}_3] = E[0] = 0$. In fact, with that constraint, $E[\hat{A}_1] = A_1 - A_3$, $E[\hat{A}_2] = A_2 - A_3$, and $E[\hat{\mu}] = \mu + A_3$. As with the $\hat{\mu} = 0$ constraint, exactly the same estimates of differences can be obtained. For example, 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 $\hat{A}_1 - \hat{A}_2$.

Finding the expectations with more complicated models is more difficult and for ease of computations would require some knowledge of matrix algebra.

Matrix algebra is also useful for ease of writing least squares and mixed model equations and properties of the mixed model procedure.

A LITTLE ABOUT MATRIX ALGEBRA

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

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

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

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

Matrix algebra is useful in working with and solving least squares and mixed model equations. 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 now.

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, let us examine the matrix by column vector multiplication:

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

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) ⟶ ⟶ ⟶ ⟶ ⇓

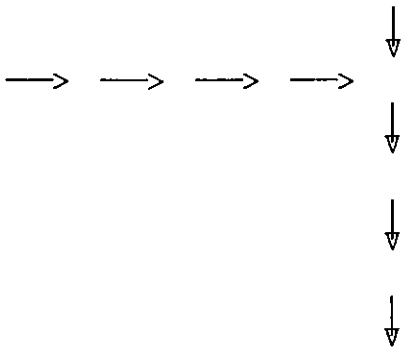
⇓

⇓

⇓

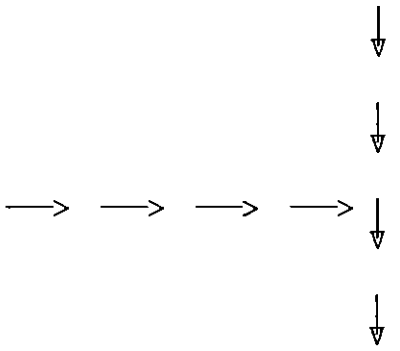
Sum for 2nd row by 1st column;

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



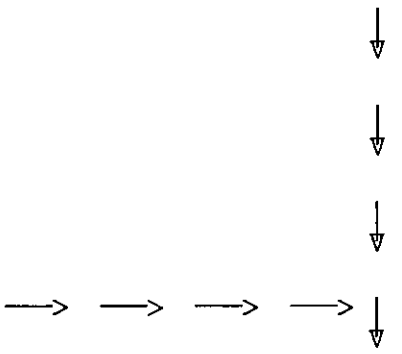
Sum for 3rd row by 1st column;

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



Sum for 4th row by 1st column;

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



The results are the left-hand sides (LHS) of the least squares equations for the 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:

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

The coefficients of the effects on the LHS of LSE or MME make up the coefficient matrix (for example, the matrix C).

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

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

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

$$Cs = r.$$

2) If C 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.

In scalar arithmetic,

$$2x = 4$$

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

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

Note that $(2)^{-1}(2) = 1$ so that $x = 2$.

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

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

If C is an invertible (i.e., nonsingular) matrix;

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

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.,

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

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 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 2x2, 3x3, 4x4, etc.

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.

3) Addition (subtraction) of two matrices is accomplished by adding (subtracting) corresponding elements of the two matrices (the matrices must have the same number of rows and columns to be conformable for addition).

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:

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

Example 2: One-way random classification model--ANIMAL MODEL

A simple one-way random classification model results when records are classified by the animal making the record. 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,

g_i is the effect on the record of the animal's genotype (usually assumed to be additive genetic effects, with $E[g_i^2] = \sigma_g^2 = h^2\sigma_y^2$, and

w_i is the residual effect of the sum of the environmental effects on y_i, $E[w_i] = 0$ 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 additive genetic values as fixed effects) and then adding σ_w^2/σ_g^2 to the diagonals of the animal (additive genetic value) equations. Let $\lambda = \sigma_w^2/\sigma_g^2 = (1-h^2)\sigma_y^2/h^2\sigma_y^2 = (1-h^2)/h^2$.

Because each animal has only one record, the mixed model equations are especially easy to write and are as follows for 3 animals:

$$\begin{aligned}
3\hat{\mu} + \hat{g}_1 + \hat{g}_2 + \hat{g}_3 &= y_. \\
\hat{\mu} + (1+\lambda)\hat{g}_1 &= y_1 \\
\hat{\mu} + (1+\lambda)\hat{g}_2 &= y_2 \\
\hat{\mu} + (1+\lambda)\hat{g}_3 &= y_3
\end{aligned}$$

where for $h^2 = .25$, $\lambda = 3$.

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 3 animal equations do not sum to the μ equation. $\hat{\mu}$ will be BLUE of μ because for this model $E[\hat{\mu}] = \mu$. $\hat{g}_1$, $\hat{g}_2$, $\hat{g}_3$ 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):

$$\hat{\mu} + (1+\lambda)\hat{g}_3 = y_3$$

$$(1+\lambda)\hat{g}_3 = y_3 - \hat{\mu} \qquad \text{(Note: } (1+\lambda) = 1 + (1-h^2)/h^2$$

$$\hat{g}_3 = h^2(y_3-\hat{\mu}) \qquad = [h^2 + (1-h^2)]/h^2$$

$$= 1/h^2 \text{)}$$

Thus $\hat{g}_3$ is the same as selection index, $I_3 = h^2(y_3-\mu)$, except that the BLUE of μ, $\hat{\mu}$, is subtracted from the animal's record rather than μ. In this example, $\hat{\mu}$ may be poorly estimated because only 3 records are used.

Note 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:

$$3\hat{\mu} + h^2(y_1 + y_2 + y_3 - 3\hat{\mu}) = y_.$$

$$3\hat{\mu} + h^2(y_. - 3\hat{\mu}) = y_.$$

$$3\hat{\mu}(1-h^2) = y_.(1-h^2) \qquad \text{and thus}$$

$$\hat{\mu} = y_./3$$

Example 3: ANIMAL MODEL with repeated records.

Although the records will be classified in one way, by animal, the effects associated with animal i on its record are of two kinds, $g_i + p_i$, 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. Again 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 animal i with $\sigma_p^2 = (r-h^2)\sigma_y^2$, [Note that $(\sigma_g^2 + \sigma_p^2)/\sigma_y^2 = r$.]

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 $\sigma_g^2 + \sigma_p^2 = r\sigma_y^2$.]

Let $\sigma_w^2/\sigma_g^2 = (1-r)/h^2 = \lambda$ 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 2 animals with n_1 and n_2 records. Thus, 5 equations will be needed corresponding to 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:

$$\begin{aligned}
n_{\cdot}\hat{\mu} + n_1\hat{g}_1 + n_2\hat{g}_2 + n_1\hat{p}_1 + n_2\hat{p}_2 &= y_{\cdot\cdot} \\
n_1\hat{\mu} + (n_1+\lambda)\hat{g}_1 + n_1\hat{p}_1 &= y_{1\cdot} \\
n_2\hat{\mu} + (n_2+\lambda)\hat{g}_2 + n_2\hat{p}_2 &= y_{2\cdot} \\
n_1\hat{\mu} + n_1\hat{g}_1 + (n_1+\gamma)\hat{p}_1 &= y_{1\cdot} \\
n_2\hat{\mu} + n_2\hat{g}_2 + (n_2+\gamma)\hat{p}_2 &= y_{2\cdot}
\end{aligned}$$

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 $E[\hat{\mu}] = \mu$. $\hat{g}_1$ and $\hat{g}_2$ will correspond to selection index for additive genetic value 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 pair of 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:

$$\lambda\hat{g}_1 = \gamma\hat{p}_1$$

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 = the sum) and with some algebra

$$\left[\frac{n_1h^2 + 1 - r + n_1r - n_1h^2}{h^2}\right]\hat{g}_1 = n_1(\bar{y}_{1.} - \hat{\mu})$$

and

$$\left[\frac{1 + (n_1-1)r}{h^2}\right]\hat{g}_1 = n_1(\bar{y}_{1.} - \hat{\mu})$$

so that

$$\hat{g}_1 = \left[\frac{n_1 h^2}{1+(n_1-1)r} \right] [\bar{y}_{1.} - \hat{\mu}]$$

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 $g_1 + p_1$:

$$\hat{g}_1 + \hat{p}_1 = \left[\frac{n_1 h^2}{1+(n_1-1)r} \right] [\bar{y}_{1.} - \hat{\mu}] + \left[\frac{r-h^2}{h^2} \right] \left[\frac{n_1 h^2}{1+(n_1-1)r} \right] [\bar{y}_{1.} - \hat{\mu}]$$

$$= \left[\frac{n_1 r}{1+(n_1-1)r} \right] [\bar{y}_{1.} - \hat{\mu}]$$

which is the selection index for producing ability with $\hat{\mu}$ instead of μ.

Example 4: One-way random classification model, SIRE MODEL

This 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 sire of the animal.

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 is equivalent to transmitting ability or 1/2 additive genetic value of the sire since a sample 1/2 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, $\sigma_w^2/\sigma_s^2 = \lambda = (1-h^2/4)/(h^2/4) = (4-h^2)/h^2$.

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 sire equations. The λ term essentially takes into account the additive relationships among animals in the same group as does the selection index procedure.

The mixed model equations become (for the case of 3 sire groups):

$$\begin{aligned}
n_. \hat{\mu} + n_1\hat{s}_1 + n_2\hat{s}_2 + n_3\hat{s}_3 &= y_{..} \\
n_1\hat{\mu} + (n_1 + \lambda)\hat{s}_1 &= y_{1.} \\
n_2\hat{\mu} + (n_2 + \lambda)\hat{s}_2 &= y_{2.} \\
n_3\hat{\mu} + (n_3 + \lambda)\hat{s}_3 &= y_{3.}
\end{aligned}$$

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:

$$\begin{aligned}
30\hat{\mu} + 11\hat{s}_1 + 4\hat{s}_2 + 15\hat{s}_3 &= y_{..} \\
11\hat{\mu} + (11 + 15)\hat{s}_1 &= y_{1.} \\
4\hat{\mu} + (4 + 15)\hat{s}_2 &= y_{2.} \\
15\hat{\mu} + (15 + 15)\hat{s}_3 &= y_{3.}
\end{aligned}$$

Note that because of the extra diagonal terms, λ, the sire equations do not add to the μ equation.

The four equations in four unknowns can be solved without imposing a constraint. $\hat{\mu}$ will be BLUE of μ since for this model $E[\hat{\mu}] = \mu$. $\hat{s}_1$, $\hat{s}_2$, $\hat{s}_3$ will be BLUP and correspond to selection indexes for transmitting ability of sires 1, 2, and 3. The equivalence to selection index can be shown by looking at any of the sire equations, e.g., sire 1, and noting that $\bar{y}_{1.} = y_{1.}/n_1$, the average of progeny of sire 1:

$$\begin{aligned}
n_1\hat{\mu} + (n_1 + 15)\hat{s}_1 &= y_{1.} \\
(n_1 + 15)\hat{s}_1 &= n_1(\bar{y}_{1.} - \hat{\mu})
\end{aligned}$$

and

$$\hat{s}_1 = \left[\frac{n_1}{n_1 + 15}\right](\bar{y}_{1.} - \hat{\mu})$$

which is the same prediction as with the selection index except that the BLUE of μ instead of μ is subtracted from the progeny average.

Example 5: Two-way fixed and random (sire) classification model

Assume the model is

$$y_{ijk} = \mu + m_i + s_j + w_{ijk}$$

where y_{ijk} is the measurement of progeny k of sire j made in management level i,

m_i is fixed effect of management i,

s_j is an effect common to animals with the same sire (j) with variance σ_s^2 = the paternal half-sib covariance = $h^2\sigma_y^2/4$, and

w_{ijk} is a random residual effect associated with progeny k of sire j in management level i, with variance $\sigma_w^2 = \sigma_y^2 - \sigma_s^2 = (1-h^2/4)\sigma_y^2)$.

Note that $\sigma_y^2 = \sigma_G^2 + \sigma_E^2$. Assume $h^2 = 1/4$. Thus, $\sigma_w^2/\sigma_s^2 = \lambda = 15$.

The following observations have been made:

$y_{111} = 530$	$y_{211} = 380$
$y_{112} = 520$	$y_{212} = 400$
$y_{121} = 460$	$y_{213} = 410$
$y_{131} = 350$	$y_{221} = 410$
$y_{132} = 340$	$y_{222} = 440$
$y_{133} = 300$	

Note that the first subscript (i) denotes the management level and the second (j) the sire of the animal. The largest third subscript (k) for a particular combination of i and j denotes the number of observations of that combination (n_{ij}).

For this example, i = 1 or 2; j = 1, 2, or 3; and $n_{11} = 2$, $n_{12} = 1$, $n_{13} = 3$, $n_{21} = 3$, $n_{22} = 2$, $n_{23} = 0$. Thus there are 6 effects in the model:

$$\mu, m_1, m_2, s_1, s_2, \text{ and } s_3 \quad .$$

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 is to make tables of the subclass numbers and sums:

n_{ij}

i \ j	1	2	3	$n_{i.}$
1	2	1	3	6
2	3	2	0	5
$n_{.j}$	5	3	3	11

$y_{ij.}$

i \ j	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:

$$\mu: \quad n_{..}\hat{\mu} + n_{1.}\hat{m}_1 + n_{2.}\hat{m}_2 + n_{.1}\hat{s}_1 + n_{.2}\hat{s}_2 + n_{.3}\hat{s}_3 = y_{...}$$

$$m_1: \quad n_{1.}\hat{\mu} + n_{1.}\hat{m}_1 + n_{11}\hat{s}_1 + n_{12}\hat{s}_2 + n_{13}\hat{s}_3 = y_{1..}$$

$$m_2: \quad n_{2.}\hat{\mu} + n_{2.}\hat{m}_2 + n_{21}\hat{s}_1 + n_{22}\hat{s}_2 + n_{23}\hat{s}_3 = y_{2..}$$

$$s_1: \quad n_{.1}\hat{\mu} + n_{11}\hat{m}_1 + n_{21}\hat{m}_2 + n_{.1}\hat{s}_1 = y_{.1.}$$

$$s_2: \quad n_{.2}\hat{\mu} + n_{12}\hat{m}_1 + n_{22}\hat{m}_2 + n_{.2}\hat{s}_2 = y_{.2.}$$

$$s_3: \quad n_{.3}\hat{\mu} + n_{13}\hat{m}_1 + n_{23}\hat{m}_2 + n_{.3}\hat{s}_3 = y_{.3.}$$

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 m_2. The number of records 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 m_1 and m_2.

To convert the least squares equations to mixed model equations, the ratio, $\sigma_w^2/\sigma_s^2 = [1 - h^2/4]\sigma_y^2/h^2\sigma_y^2/4 = 15$, is added to the diagonal coefficients 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. Let $\hat{\mu} = 0$ and eliminate that equation. The equations to solve become:

$$
\begin{aligned}
6\hat{m}_1 + 2\hat{s}_1 + \hat{s}_2 + 3\hat{s}_3 &= 2500\\
5\hat{m}_2 + 3\hat{s}_1 + 2\hat{s}_2 &= 2040\\
2\hat{m}_1 + 3\hat{m}_2 + 20\hat{s}_1 &= 2240\\
\hat{m}_1 + 2\hat{m}_2 + 18\hat{s}_2 &= 1310\\
3\hat{m}_1 + 18\hat{s}_3 &= 990
\end{aligned}
$$

Solutions are:

$$\hat{\mu} = 0 \qquad \hat{s}_1 = 10$$

$$\hat{m}_1 = 420 \qquad \hat{s}_2 = 5$$

$$\hat{m}_2 = 400 \qquad \hat{s}_3 = -15$$

Note that $\hat{s}_1 + \hat{s}_2 + \hat{s}_3 = 0$. This property holds for any classification of random effects. Note that 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 course, 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 $E[\hat{m}_2] = \mu + m_2$ with the constraint $\hat{\mu} = 0$.

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,

no constraints would be imposed, $\hat{m}_1 + \hat{m}_2 = 0$, and

$E[\hat{\mu}] = \mu$ but only when no other fixed effects are in the model.

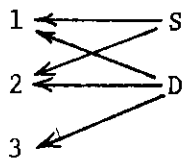
Note that then $\sigma_y^2 = \sigma_m^2 + \sigma_s^2 + \sigma_w^2$ so that σ_w^2/σ_s^2 may be different from when management levels are considered to be fixed effects.

Example 6: The ANIMAL MODEL with animals related

The selection index takes advantage of records of relatives to improve predictions. The mixed model procedure can as well. Instead of adding σ_w^2/σ_g^2 to the least squares diagonal 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. The least squares equations are:

$$\left(\begin{array}{c|ccc} 3 & 1 & 1 & 1 \\ \hline 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{array}\right) \begin{pmatrix} \hat{\mu} \\ \hat{g}_1 \\ \hat{g}_2 \\ \hat{g}_3 \end{pmatrix} = \begin{pmatrix} y_. \\ y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

To convert the least squares equations to mixed model equations, λA^{-1} is added to the block of coefficients for the g equations (outlined in broken lines). For example, if $\sigma_w^2/\sigma_g^2 = (1-h^2)/h^2 = 3$, then

$$\lambda A^{-1} = 3 \begin{bmatrix} 15/11 & -7/11 & -2/11 \\ -7/11 & 15/11 & -2/11 \\ -2/11 & -2/11 & 12/11 \end{bmatrix}$$

and the mixed model equations become:

$$\begin{bmatrix} 3 & 1 & 1 & 1 \\ 1 & 1 + 45/11 & 0 - 21/11 & 0 - 6/11 \\ 1 & 0 - 21/11 & 1 + 45/11 & 0 - 6/11 \\ 1 & 0 - 6/11 & 0 - 6/11 & 1 + 36/11 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{g}_1 \\ \hat{g}_2 \\ \hat{g}_3 \end{bmatrix} = \begin{bmatrix} y_. \\ y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

Example 7: ANIMAL MODEL with 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 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 Rules for Calculation of A^{-1}.) The method, however, requires including 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}. The mixed model equations for animals with records must be augmented with equations for the ancestors without records. Let A_+^{-1} be the inverse of A_+ which includes the ancestors without records that create relationships among the animals with records. The coefficients of the least squares equations for animals without records are all zero and the right-hand sides are also zero. When λA_+^{-1} is added to the block of coefficients for the animals including the ancestors without records, then the coefficients are not all zero for the ancestor equations although the right-hand sides are zero.

The procedure will be illustrated with the previous example. The least squares equations are:

$$\left[\begin{array}{cc|cccc} 3 & 1 & 1 & 1 & 0 & 0 \\ \hline 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \left[\begin{array}{c} \hat{\mu} \\ \hat{g}_1 \\ \hat{g}_2 \\ \hat{g}_3 \\ \hat{g}_S \\ \hat{g}_D \end{array}\right] = \left[\begin{array}{c} y_. \\ y_1 \\ y_2 \\ y_3 \\ 0 \\ 0 \end{array}\right]$$

Using the rules for calculating A_+^{-1} with S and D included at the end to agree with the order in the least squares equations:

$$A_+^{-1} = \begin{pmatrix} 2 & 0 & 0 & -1 & -1 \\ 0 & 2 & 0 & -1 & -1 \\ 0 & 0 & 4/3 & 0 & -2/3 \\ -1 & -1 & 0 & 2 & 1 \\ -1 & -1 & -2/3 & 1 & 2+1/3 \end{pmatrix}$$

Then λA_+^{-1} is added to the block of coefficients outlined with interrupted lines.

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} was added to equations for g_1, g_2, and g_3 in the previous example. 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.

A simpler example will illustrate. 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 $\lambda = \sigma_w^2/\sigma_g^2 = (1-r)/h^2$ and $\gamma = \sigma_w^2/\sigma_p^2 = (1-r)/(r-h^2)$, the augmented equations are:

$$\begin{bmatrix} n & n & n & 0 & 0 \\ n & n+\lambda & n & 0 & 0 \\ n & n & n+2\lambda & -\lambda & -\lambda \\ 0 & 0 & -\lambda & 3\lambda/2 & \lambda/2 \\ 0 & 0 & -\lambda & \lambda/2 & 3\lambda/2 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{p}_C \\ \hat{g}_C \\ \hat{g}_S \\ \hat{g}_D \end{bmatrix} = \begin{bmatrix} y_{1.} \\ y_{1.} \\ y_{1.} \\ 0 \\ 0 \end{bmatrix}$$

Note that $\hat{g}_S = \hat{g}_D$ because each has the same relationship to their progeny which has the 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:

$$n\hat{\mu} + n\hat{p}_C + (n+2\lambda)\hat{g}_C - (\lambda/2)\hat{g}_C - (\lambda/2)\hat{g}_C = y_{1.}$$

This equation, on combining terms, is the same as the equation for g_C if relationships to S and D had been ignored:

$$n\hat{\mu} + n\hat{p}_C + (n+\lambda)\hat{g}_C = y_{1.}$$

Such a result is expected because neither S nor D contributed any 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 easier than calculating A and then A^{-1} for animals with records.

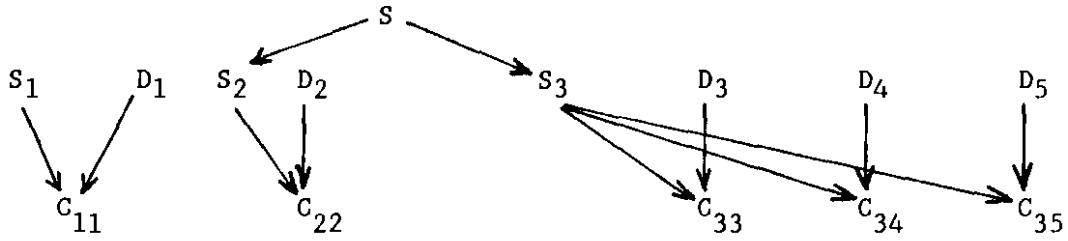
This example suggests that an animal model can be used to evaluate sires that have many progeny. In the simple example the evaluation of S was obtained and based on only one progeny.

Example 8: 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 has a son that has progeny but the sire of the son has no progeny with records nor any other sons with progeny or descendants in the group of animals to be evaluated.

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 but the solutions will be the same and more equations would have 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, 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 $\sigma_w^2/\sigma_g^2 = (1-h^2)/h^2 = \delta$ are:

$$
\begin{pmatrix}
10 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
 & 1+3\delta/2 & 0 & 0 & 0 & 0 & -\delta & 0 & 0 & 0 & 0 & 0 & \delta/2 & 0 & 0 \\
 & & 1+3\delta/2 & 0 & 0 & 0 & 0 & -\delta & 0 & 0 & 0 & 0 & 0 & \delta/2 & 0 \\
 & & & 1+3\delta/2 & 0 & 0 & 0 & 0 & -\delta & 0 & 0 & 0 & 0 & 0 & \delta/2 \\
 & & & & 1+3\delta/2 & 0 & 0 & 0 & 0 & -\delta & 0 & 0 & 0 & 0 & \delta/2 \\
 & & & & & 1+3\delta/2 & 0 & 0 & 0 & 0 & -\delta & 0 & 0 & 0 & \delta/2 \\
 & & & & & & 1+2\delta & 0 & 0 & 0 & 0 & 0 & -\delta & 0 & 0 \\
 & & & & & & & 1+2\delta & 0 & 0 & 0 & 0 & 0 & -\delta & 0 \\
 & & & & & & & & 1+2\delta & 0 & 0 & 0 & 0 & 0 & -\delta \\
 & & & & & & & & & 1+2\delta & 0 & 0 & 0 & 0 & -\delta \\
 & & & & & & & & & & 1+2\delta & 0 & 0 & 0 & -\delta \\
 & \text{SYMMETRIC} & & & & & & & & & & 5\delta/3 & 0 & -2\delta/3 & -2\delta/3 \\
 & \text{(coefficients in a column below the} & & & & & & & & & & & 3\delta/2 & 0 & 0 \\
 & \text{diagonal are the same as those in} & & & & & & & & & & & & 11\delta/6 & 0 \\
 & \text{the row to the right of the diagonal)} & & & & & & & & & & & & & 17\delta/6
\end{pmatrix}
\begin{pmatrix}
\hat{\mu} \\
\hat{g}_{D_1} \\
\hat{g}_{D_2} \\
\hat{g}_{D_3} \\
\hat{g}_{D_4} \\
\hat{g}_{D_5} \\
\hat{g}_{C_{11}} \\
\hat{g}_{C_{22}} \\
\hat{g}_{C_{33}} \\
\hat{g}_{C_{34}} \\
\hat{g}_{C_{35}} \\
\hat{g}_{S} \\
\hat{g}_{S_1} \\
\hat{g}_{S_2} \\
\hat{g}_{S_3}
\end{pmatrix}
=
\begin{pmatrix}
y_{\cdot} \\
y_{D_1} \\
y_{D_2} \\
y_{D_3} \\
y_{D_4} \\
y_{D_5} \\
y_{C_{11}} \\
y_{C_{22}} \\
y_{C_{33}} \\
y_{C_{34}} \\
y_{C_{35}} \\
0 \\
0 \\
0 \\
0
\end{pmatrix}
$$

Examination of the solution for a sire shows the weight for each mate is -1/2 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 leave the part of the progeny's genetic value contributed by the sire. For example, for S_3:

$$(17\delta/6)\hat{g}_{S_3} = (2\delta/3)\hat{g}_S + \delta(\hat{g}_{C_{33}} + \hat{g}_{C_{34}} + \hat{g}_{C_{35}}) - (\delta/2)(\hat{g}_{D_3} + \hat{g}_{D_4} + \hat{g}_{D_5})$$

$$(17\delta/6)\hat{g}_{S_3} = (2\delta/3)\hat{g}_S + \delta[(\hat{g}_{C_{33}} - \hat{g}_{D_3}/2) + (\hat{g}_{C_{34}} - \hat{g}_{D_4}/2) + (\hat{g}_{C_{35}} - \hat{g}_{D_5}/2)]$$

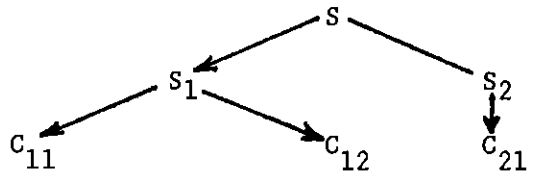
And for S_1:

$$\hat{g}_{S_1} = (2\delta/3)(\hat{g}_{C_{11}} - \hat{g}_{D_1}/2)$$

Example 9: Sire evaluation with ANIMAL MODEL ignoring mates and relationships through females

In the past, sire evaluations generally have ignored 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 (note 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.

Example 10: Sire evaluation with SIRE MODEL ignoring mates and female relationships

The approximate animal model described in the previous example requires an equation for each progeny which with many animals may be computationally prohibitive. The number of equations can be reduced essentially to the number of sires by using the sire model. Only male to male relationships will be considered (assumes dams unrelated to sires and to each other). The sire model is:

$$y_{ij} = \mu + s_i + w_{ij}$$

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 will be used. Now only relationships among S_1, S_2, and S need to 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

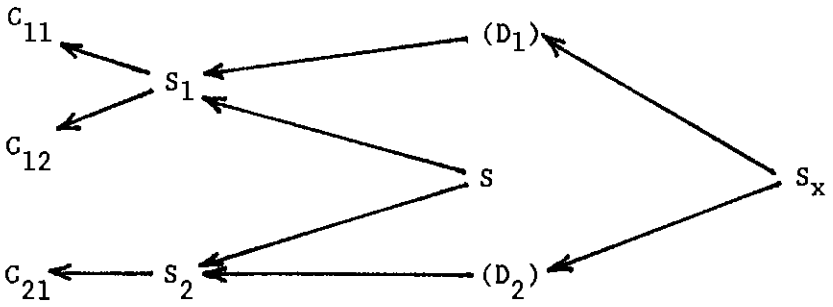
augmented equations for the sires that had no records themselves. That $\hat{s}_1 = \hat{g}_{S_1}/2$ can be shown by absorbing equations for g_{11} and g_{12} in example 9 into the equation for g_{S_1}. The equation for g_{S_1} will be the same as for s_1 except that the coefficients for $\hat{g}_{S_1}$ 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.

Example 11: Sire evaluation with SIRE MODEL ignoring female relationships but calculating A_+^{-1} 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 and if they have no progeny with records can be included as 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 the section on 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:



D_1 and D_2 will not be included in A_+^{-1}. With males ordered S_1, S_2, S, S_x:

$$A_+^{-1} = (1/11) \begin{pmatrix} 16 & 0 & -8 & -4 \\ 0 & 16 & -8 & -4 \\ -8 & -8 & 19 & 4 \\ -4 & -4 & 4 & 13 \end{pmatrix}$$

As 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 augmented equations for S and S_x. The augmented mixed model equations are:

$$\begin{bmatrix} 3 & 2 & 1 & 0 & 0 \\ & 2+16\gamma/11 & 0 & -8\gamma/11 & -4\gamma/11 \\ & & 1+16\gamma/11 & -8\gamma/11 & -4\gamma/11 \\ & & & 19\gamma/11 & 4\gamma/11 \\ \text{Symmetric} & & & & 13\gamma/11 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_S \\ \hat{s}_x \end{bmatrix} = \begin{bmatrix} y_{..} \\ y_{1.} \\ y_{2.} \\ 0 \\ 0 \end{bmatrix}$$

The variance of prediction errors, e.g., $V(\hat{s}_i - s_i)$, will be decreased as compared to ignoring S_x.

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. How to calculate them will be described for T in the model: genetic value, T = g; transmitting ability, T = s; producing ability, T = g + p; as well as for fixed effects estimated from data. The first calculation is for $V(T-\hat{T})$ which 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 $\gamma = \sigma^2_w/\sigma^2_p$ and the symbolic mixed model equations be:

$$\begin{pmatrix} c_{\mu,\mu} & c_{\mu,p_1} & \cdots & c_{\mu,g_1} & \cdots \\ c_{p_1,\mu} & c_{p_1,p_1} & \cdots & c_{p_1,g_1} & \cdots \\ \vdots & & & & \\ c_{g_1,\mu} & c_{g_1,p_1} & \cdots & c_{g_1,g_1} & \cdots \\ \vdots & & & & \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{p}_1 \\ \vdots \\ \hat{g}_1 \\ \vdots \end{pmatrix} = \begin{pmatrix} y_{..} \\ y_{1.} \\ \vdots \\ y_{1.} \\ \vdots \end{pmatrix}$$

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 =

$$\begin{pmatrix} n_1 & 0 & \cdots \\ 0 & n_2 & \\ \vdots & & \ddots \end{pmatrix} + \lambda A^{-1}$$

In matrix terms

$$Cs = r$$

where the solution vector:

$$s = C^{-1}r$$

Let the terms of the inverse of C be

$$C^{-1} = \begin{pmatrix} c^{\mu,\mu} & c^{\mu,p1} & \cdots & c^{\mu,g1} & \cdots \\ c^{p1,\mu} & c^{p1,p1} & \cdots & c^{p1,g1} & \cdots \\ \vdots & & & & \\ c^{g1,\mu} & c^{g1,p1} & \cdots & c^{g1,g1} & \cdots \\ \vdots & & & & \end{pmatrix}$$

These terms when multiplied by σ_w^2 correspond to prediction error variances and covariances:

$$V(g_1-\hat{g}_1) = c^{g1,g1}\sigma_w^2$$

$$V(p_1-\hat{p}_1) = c^{p1,p1}\sigma_w^2$$

$$Cov(g_1-\hat{g}_1,\ p_1-\hat{p}_1) = c^{g1,p1}\sigma_w^2$$

so that for producing ability;

$$V[(g_1+p_1) - (\hat{g}_1+\hat{p}_1)] = V(g_1-\hat{g}_1) + V(p_1-\hat{p}_1) + 2cov(g_1-\hat{g}_1,\ p_1-\hat{p}_1)$$
$$= (c^{g1,g1} + c^{p1,p1} + 2c^{g1,p1})\sigma_w^2$$

Variance of prediction error of genetic value:

Because $V(g_i-\hat{g}_i) = c^{g_i,g_i}\sigma_w^2$ and also $V(g_i-\hat{g}_i) = (1-r^2_{g\hat{g}})\sigma_g^2$, then $c^{g_i,g_i}\sigma_w^2 = (1-r^2_{g\hat{g}})\sigma_g^2$. The equation can be solved for $r^{2\wedge}_{gg}$:

$$r^{2\wedge}_{gg} = 1 - c^{g_i,g_i}(\sigma_w^2/\sigma_g^2) \quad .$$

The ratio $\sigma_w^2/\sigma_g^2 = (1-r)/h^2$ must be known and c^{g_i,g_i} is from the appropriate element of the inverse of C.

Variance of prediction error of producing ability

Because $V[(g_i+p_i) - (\hat{g}_i+\hat{p}_i)] = (c^{g_i,g_i} + c^{p_i,p_i} + 2c^{g_i,p_i})\sigma_w^2$ and also $= (1-r^2_{g+p,\hat{g}+\hat{p}})(\sigma_g^2 + \sigma_p^2)$, then $r^2_{g+p,\hat{g}+\hat{p}} = 1 - (c^{g_i,g_i} + c^{p_i,p_i} + 2c^{g_i,p_i})\sigma_w^2/(\sigma_g^2+\sigma_p^2)$. The ratio $\sigma_w^2/(\sigma_g^2+\sigma_p^2) = (1-r)/r$ must be known as well as elements from C^{-1}.

Variance of prediction error for other models

For other models calculations are similar to those described in previous paragraphs. The appropriate elements from 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 C^{-1} corresponding to the random effects, g, p, s, etc., are used for calculating variances and covariances of prediction errors for those effects.

These elements, however, will depend 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 $\overset{*}{C}$ is nonsingular, then $\overset{*}{C}^{-1}$ exists 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, g, p, s, will be the same for any set of permissable constraints. Similarly, prediction error variances for the random effects do not depend on the constraints chosen, i.e., the block of elements of $\overset{*}{C}^{-1}$ corresponding to the random effects are unique and do 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 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, $E[\hat{f}_i - \hat{f}_j]$ may equal $f_i - f_j$ for levels i and j of fixed factor f. The variance of the estimable difference, $\hat{f}_i - \hat{f}_j$ is $V(\hat{f}_i - \hat{f}_j) = (c^{f_i,f_i} + c^{f_j,f_j} - 2c^{f_i,f_j})\sigma_w^2$.

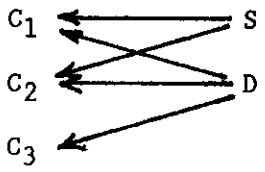
Numerical example of ANIMAL MODEL with a fixed classification and augmented equations with different constraints

The model for a record k of cow j affected by level i of fixed factor f is:

$$y_{ijk} = \mu + f_i + p_j + g_j + w_{ijk}$$

Let $\sigma_w^2 = (2000)^2$ and with $r = .6$ and $h^2 = .4$, $\sigma_w^2/\sigma_g^2 = (1-r)/h^2 = \lambda = 1$ and $\sigma_w^2/\sigma_p^2 = (1-r)/(r-h^2) = 2$.

The cows 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 Example 7: ANIMAL MODEL with equations augmented for relatives without records.)



The records for the animals are distributed in the levels of the fixed factor as follows:

Fixed factor	Cow C_1	Cow C_2	Cow C_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
Cow totals	22,000	15,000	31,000	68,000

The augmented mixed model equations are:

$$
\begin{pmatrix}
6 & 2 & 2 & 2 & 2 & 1 & 3 & 2 & 1 & 3 & 0 & 0 \\
 & 2 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\
 & & 2 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\
 & & & 2 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\
 & & & & 2+2 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\
 & & & & & 1+2 & 0 & 0 & 1 & 0 & 0 & 0 \\
 & & & & & & 3+2 & 0 & 0 & 3 & 0 & 0 \\
 & & & & & & & 2+2 & 0 & 0 & -1 & -1 \\
 & & & & & & & & 1+2 & 0 & -1 & -1 \\
 & & & & & & & & & 3+(4/3) & 0 & -2/3 \\
 & \text{SYMMETRIC} & & & & & & & & & +2 & +1 \\
 & & & & & & & & & & & +2(1/3)
\end{pmatrix}
\begin{pmatrix} \hat{\mu} \\ \hat{f}_1 \\ \hat{f}_2 \\ \hat{f}_3 \\ \hat{p}_1 \\ \hat{p}_2 \\ \hat{p}_3 \\ \hat{g}_1 \\ \hat{g}_2 \\ \hat{g}_3 \\ \hat{g}_S \\ \hat{g}_D \end{pmatrix}
=
\begin{pmatrix} 68{,}000 \\ 19{,}000 \\ 22{,}000 \\ 27{,}000 \\ 22{,}000 \\ 15{,}000 \\ 31{,}000 \\ 22{,}000 \\ 15{,}000 \\ 31{,}000 \\ 0 \\ 0 \end{pmatrix}
$$

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:

$$\begin{pmatrix}
1.310 & .000 & -.500 & -.657 & -.222 & -.065 & -.213 & -.616 & -.458 & -.569 & -.287 & -.500 \\
.000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 \\
-.500 & .000 & 1.000 & .500 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 \\
-.657 & .000 & .500 & 1.296 & .139 & -.157 & .019 & .130 & -.167 & .028 & -.019 & .000 \\
-.222 & .000 & .000 & .139 & .417 & .028 & .056 & -.111 & .000 & .083 & -.056 & .000 \\
-.065 & .000 & .000 & -.157 & .028 & .435 & .037 & .009 & -.083 & .056 & -.037 & .000 \\
-.213 & .000 & .000 & .019 & .056 & .037 & .407 & .102 & .083 & -.139 & .093 & .000 \\
-.616 & .000 & .000 & .130 & -.111 & .009 & .102 & .838 & .458 & .403 & .398 & .500 \\
-.458 & .000 & .000 & -.167 & .000 & -.083 & .083 & .458 & .875 & .375 & .417 & .500 \\
-.569 & .000 & .000 & .028 & .083 & .056 & -.139 & .403 & .375 & .792 & .139 & .500 \\
-.287 & .000 & .000 & -.019 & -.056 & -.037 & .093 & .398 & .417 & .139 & .907 & .000 \\
-.500 & .000 & .000 & .000 & .000 & .000 & .000 & .500 & .500 & .500 & .000 & 1.000
\end{pmatrix}$$

With the constraint $\hat{\mu} = 0$, coefficients for the μ row and column are zeroed.

The inverse is:

$$\begin{pmatrix}
.000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 \\
.000 & 1.310 & .810 & .653 & -.222 & -.065 & -.213 & -.616 & -.458 & -.569 & -.287 & -.500 \\
.000 & .810 & 1.310 & .653 & -.222 & -.065 & -.213 & -.616 & -.458 & -.569 & -.287 & -.500 \\
.000 & .653 & .653 & 1.292 & -.083 & -.222 & -.194 & -.486 & -.625 & -.542 & -.306 & -.500 \\
.000 & -.222 & -.222 & -.083 & .417 & .028 & .056 & -.111 & .000 & .083 & -.056 & .000 \\
.000 & -.065 & -.065 & -.222 & .028 & .435 & .037 & .009 & -.083 & .056 & -.037 & .000 \\
.000 & -.213 & -.213 & -.194 & .056 & .037 & .407 & .102 & .083 & -.139 & .093 & .000 \\
.000 & -.616 & -.616 & -.486 & -.000 & .009 & .102 & .838 & .458 & .403 & .398 & .500 \\
.000 & -.458 & -.458 & -.625 & .000 & -.083 & .083 & .458 & .875 & .375 & .417 & .500 \\
.000 & -.569 & -.569 & -.542 & .083 & .056 & -.139 & .403 & .375 & .792 & .139 & .500 \\
.000 & -.287 & -.287 & -.306 & -.056 & -.037 & .093 & .398 & .417 & .139 & .907 & .000 \\
.000 & -.500 & -.500 & -.500 & .000 & .000 & .000 & .500 & .500 & .500 & .000 & 1.000
\end{pmatrix}$$

With the constraint $\hat{f}_3 = 0$ the inverse is:

$$\begin{pmatrix}
1.292 & -.639 & -.639 & .000 & -.083 & -.222 & -.194 & -.486 & -.625 & -.542 & -.306 & -.500 \\
-.639 & 1.296 & .796 & .000 & -.139 & .157 & -.019 & -.130 & .167 & -.028 & .019 & .000 \\
-.639 & .796 & 1.296 & .000 & -.139 & .157 & -.019 & -.130 & .167 & -.028 & .019 & .000 \\
.000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 \\
-.083 & -.139 & -.139 & .000 & .417 & .028 & .056 & -.111 & .000 & .083 & -.056 & .000 \\
-.222 & .157 & .157 & .000 & .028 & .435 & .037 & .009 & -.083 & .056 & -.037 & .000 \\
-.194 & -.019 & -.019 & .000 & .056 & .037 & .407 & .102 & .083 & -.139 & .093 & .000 \\
-.486 & -.130 & -.130 & .000 & -.111 & .009 & .102 & .838 & .458 & .403 & .398 & .500 \\
-.625 & .167 & .167 & .000 & .000 & -.083 & .083 & .458 & .875 & .375 & .417 & .500 \\
-.542 & -.028 & -.028 & .000 & .083 & .056 & -.139 & .403 & .375 & .792 & .139 & .500 \\
-.306 & .019 & .019 & .000 & -.056 & -.037 & .093 & .398 & .417 & .139 & .907 & .000 \\
-.500 & .000 & .000 & .000 & .000 & .000 & .000 & .500 & .500 & .500 & .000 & 1.000
\end{pmatrix}$$

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_D and g_S are the same. The solutions for those effects are also the same as is shown in the table. 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:

$$V(g_1-\hat{g}_1) = c^{g_1,g_1}\sigma_w^2 = .838(2000)^2 \quad .$$

$$V(g_D-\hat{g}_D) = c^{g_D,g_D}\sigma_w^2 = 1.000(2000)^2 \quad .$$

$$V[(g_1+p_1) - (\hat{g}_1+\hat{p}_1)] = c^{p_1,p_1} + c^{g_1,g_1} + 2c^{p_1,g_1})\sigma_w^2$$

$$= [.417 + .838 + 2(-.111)](2000)^2 \quad .$$

Because what $\hat{f}_i$ estimates depends on the constraint, $V(\hat{f}_i)$ is different from constraint to constraint.

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 the table 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 was used to obtain a set of solutions.

Solutions for augmented and nonagumented mixed model equations with different constraints

	Augmented			Nonaugmented
Solution	$\hat{f}_1=0$	$\hat{\mu}=0$	$\hat{f}_3=0$	$\hat{f}_3=0$
$\hat{\mu}$	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	--

If the mixed model equations had not been augmented but S and D had been used to calculate 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 mixed model equations are:

$$\begin{pmatrix}
6 & 2 & 2 & 2 & 2 & 1 & 3 & 2 & 1 & 3 \\
 & 2 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\
 & & 2 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\
 & & & 2 & 0 & 1 & 1 & 0 & 1 & 1 \\
 & & & & 4 & 0 & 0 & 2 & 0 & 0 \\
 & & & & & 3 & 0 & 0 & 1 & 0 \\
 & & & & & & 5 & 0 & 0 & 3 \\
 & & & & & & & 3.364 & -.636 & -.182 \\
 & & & & & & & & 2.364 & -.182 \\
 & \text{SYMMETRIC} & & & & & & & & 4.091
\end{pmatrix}
\begin{pmatrix} \hat{\mu} \\ \hat{f}_1 \\ \hat{f}_2 \\ \hat{f}_3 \\ \hat{p}_1 \\ \hat{p}_2 \\ \hat{p}_3 \\ \hat{g}_1 \\ \hat{g}_2 \\ \hat{g}_3 \end{pmatrix}
=
\begin{pmatrix} 68{,}000 \\ 19{,}000 \\ 22{,}000 \\ 27{,}000 \\ 22{,}000 \\ 15{,}000 \\ 31{,}000 \\ 22{,}000 \\ 15{,}000 \\ 31{,}000 \end{pmatrix}$$

The solutions shown in the table 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 $\hat{f}_3 = 0$.

$$\begin{pmatrix}
1.292 & -.639 & -.639 & .000 & -.083 & -.222 & -.194 & -.486 & -.625 & -.542 & .000 & .000 \\
-.639 & 1.296 & .796 & .000 & -.139 & .157 & -.019 & -.130 & .167 & -.028 & .000 & .000 \\
-.639 & .796 & 1.296 & .000 & -.139 & .157 & -.019 & -.130 & .167 & -.028 & .000 & .000 \\
.000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 \\
-.083 & -.139 & -.139 & .000 & .417 & .028 & .056 & -.111 & .000 & .083 & .000 & .000 \\
-.222 & .157 & .157 & .000 & .028 & .435 & .037 & .009 & -.083 & .056 & .000 & .000 \\
-.194 & -.019 & -.019 & .000 & .056 & .037 & .407 & .102 & .083 & -.139 & .000 & .000 \\
-.486 & -.130 & -.130 & .000 & -.111 & .009 & .102 & .838 & .458 & .403 & .000 & .000 \\
-.625 & .167 & .167 & .000 & .000 & -.083 & .083 & .458 & .875 & .375 & .000 & .000 \\
-.542 & -.028 & -.028 & .000 & .083 & .056 & -.139 & .403 & .375 & .792 & .000 & .000 \\
.000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 \\
.000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000 & .000
\end{pmatrix}$$

SOLVING LEAST SQUARES AND MIXED MODEL EQUATIONS

In many situations, the number of equations is so large that an inverse of the coefficient matrix cannot be computed. 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.

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:

$$c_{11}s_1 + c_{12}s_2 + c_{13}s_3 = r_1$$
$$c_{21}s_1 + c_{22}s_2 + c_{23}s_3 = r_2$$
$$c_{31}s_1 + c_{32}s_2 + c_{33}s_3 = r_3$$

1) To start 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.

2) The basic step for each equation is to solve for that solution after substituting solutions from same or previous rounds of iteration for the other solutions.

Round 1

i) Solve for s_1 with s_2^o and s_3^o:

$$s_1^1 = (1/c_{11})[r_1 - c_{12}s_2^o - c_{13}s_3^o]$$

Replace the previous solution for s_1 with s_1^1.

ii) Solve for s_2 with s_1^1 and s_3^0:

$$s_2^1 = (1/c_{22})[r_2 - c_{21}s_1^1 - c_{23}s_3^0]$$

Replace the previous solution for s_2 with s_2^1.

iii) Solve for s_3 with s_1^1 and s_2^1:

$$s_3^1 = (1/c_{33})[r_3 - c_{31}s_1^1 - c_{32}s_2^1]$$

Replace the previous solution for s_3 with s_3^1.

Round 2 $\longrightarrow$ n

i) $s_1^n \longleftarrow (1/c_{11})[r_1 - c_{12}s_2^{n-1} - c_{13}s_3^{n-1}]$

ii) $s_2^n \longleftarrow (1/c_{22})[r_2 - c_{21}s_1^n - c_{23}s_3^{n-1}]$

iii) $s_3^n \longleftarrow (1/c_{33})[r_3 - c_{31}s_1^n - c_{32}s_2^n]$

Note that the most current estimates in s are used. The Jacobi method does not update s until at the end of the round.

An equivalent expression for s_i^n is

$$s_i^n = s_i^{n-1} + (1/c_{ii})[r_i - \sum_{j=1}^{i-1} c_{ij}s_j^n - c_{ii}s_i^{n-1} - \sum_{j=i+1} c_{ij}s_j^{n-1}]$$

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 brackets on the right which would be zero when the solutions are correct. Thus, the 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 $(\Sigma e_i^2)^{1/2} / (\Sigma r_i^2)^{1/2}$ 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 $(\Sigma r_i^2)^{1/2}$ 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 is that a modification of Gauss-Seidel iteration called successive-over-relaxation is easy to implement:

$$s_i^n = s_i^{n-1} + (w/c_{ii})[r_i - \sum_{j=1}^{i-1} c_{ij}s_j^n - c_{ii}s_i^{n-1} - \sum_{j=1+1} c_{ij}s_i^{n-1}]$$

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). The difficulty is to find the optimum w.

COMPUTING ALGORITHM FOR LEAST SQUARE EQUATIONS

Computing strategies for accumulating the coefficients and right-hand sides of the least squares equations naturally depend on the amount of data, the model, and computer memory. Nevertheless, a symbolic algorithm can be used to remember what coefficients are involved for each record. Data usually are presented for computing one record at a time. Coefficients and right-hand sides for each record are summed into computer memory 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. The record carries four elements of the model (excluding w_{ijk}) to each sum. Thus, each record contributes to 16 elements of the coefficient matrix. 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 elements in C where a 1 will be added for that record:

$$\mu\mu,\ \mu f_i,\ \mu p_j,\ \mu g_j\ ,$$

$$f_i\mu,\ f_i f_i,\ f_i p_j,\ f_i g_j\ ,$$

$$p_j\mu,\ p_j f_i,\ p_j p_j,\ p_j g_j\ ,$$

$$g_j\mu,\ g_j f_i,\ g_j p_j,\ g_j g_j\ .$$

For example, the diagonals $c_{\mu\mu}$, $c_{f_i f_i}$, $c_{p_j p_j}$, $c_{g_j g_j}$ will add a 1 and the offdiagonal coefficients represented by products such as μf_i will also add a 1. Notice that 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.

Flow chart for mixed model equations (MME)

I. Model

A. Fixed factors

B. Random factors (other than G, G/2, PE), $_r$

C. G + E; $\sigma_g^2 + \sigma_e^2 = \sigma_p^2$ ($\sigma_y^2 = \sigma_p^2$ + sum of σ_r^2's)

(animal model) 1. G; W = E, $\sigma_g^2 = h^2\sigma_p^2$, $\sigma_w^2 = (1-h^2)\sigma_p^2$

(repeated records) 2. G, PE; W = TE, $\sigma_g^2 = h^2\sigma_p^2$, $\sigma_{pe}^2 = (r-h^2)\sigma^2$, $\sigma_w^2 = (1-r)\sigma_p^2$

(sire model) 3. 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. L.S.E. (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, g equations

2. If multiple records, add σ_w^2/σ_{pe}^2, diagonals of pe equations

C. Sires with progeny with records, unrelated sires

1. Add σ_w^2/σ_s^2 to diagonal coefficients, s equations

IV. Modifications (animals or sires related)

A. Add σ_w^2/σ_r^2 to diagonal coefficients, other random factors.

B. Animals with records related, A is table of relationships.

1. Multiple records, add σ_w^2/σ_{pe}^2 to diagonal coefficients of pe equations.

2. Add A^{-1} (σ_w^2/σ_g^2) to g x g block of coefficients.

OR

3. Calculate A_+^{-1} directly by 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 having 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 (noninbred) (rules for 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)

   c) Jointly predict $G/2$ for sires with progeny records and relatives with no progeny records.

Summary XVII

COMPUTING THE INVERSE OF THE RELATIONSHIP MATRIX

Mixed model procedures require the inverse of the relationship matrix when genetic values are to be predicted (BLUP) to account for the covariances among the genetic effects in the model. The usual procedure would be to compute the relationship matrix (Table) and have a computer program find the inverse. Henderson (1976) has found, however, a rapid way of calculating the inverse of the relationship matrix directly. If the animals are non-inbred or assumed to be non-inbred, the procedure is very rapid. Ignoring a small amount of inbreeding probably is a good approximation in most prediction problems.

The computing steps involve adding from one to six values to different parts of the inverse of the relationship matrix for each animal depending on how many parents are known. After all animals have been processed, the result is the inverse of the relationship matrix. The inverse elements each multiplied by the proper variance ratio, σ_w^2/σ_v^2, are used to modify the least squares equations to make the mixed model equations.

The animals can be processed in any order. Base animals must be included even though they may not have records. (Base animals will be those which establish relationships among other animals but are not themselves related.)

Because the base animals may not have records the mixed model equations are expanded as was illustrated in the cow evaluation model to include an equation for each base animal with a zero sum on the right-hand

side and coefficients on the left-hand side made up of inverse elements of the relationship matrix multiplied by σ_w^2/σ_v^2.

Any base animal with only one relative [an animal with records (sire with progeny with records for the sire model) or another base animal] need not be included in the expanded mixed model equations. Thus, such a base animal need not be included in the inverse of the relationship matrix and can be listed as unknown. If such a base animal is included in the inverse of the relationship matrix, then an equation for that animal must be added to the mixed model equations as for any other base animal.

Because the computing procedure can accept animals in any order, putting the base animals at the end of the inverse table (in contrast to the usual way of computing relationships) may make setting up the equations easier.

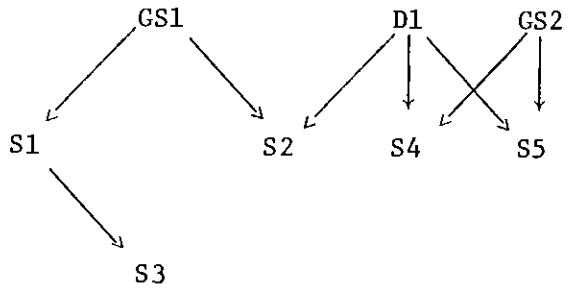
The simple rules for building the inverse of the relationship matrix for non-inbred animals are:

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: for example, if -2/3 to (s,p), then -2/3 to (p,s).

Note that p, s, and d will be animal numbers, and (p,p), (s,p), etc., combinations refer to a location in the inverse table.

In the example that follows, three animals are base animals: GS1, D1, and GS2. The other five animals are related through them. GS1, D1, and GS2 must be included in building the table even though predictions are wanted only for S1,...,S5. The relationships for the example are diagrammed below.



The table will be built beginning with the information for S1; a blank indicates the parent is not known. Actually if an animal is a parent of only one animal then that parent does not need to be included as a base animal.

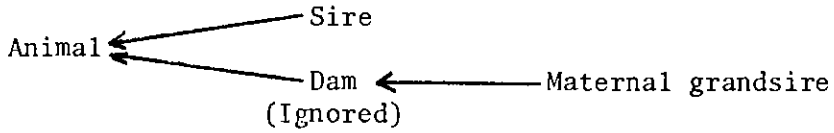
p	s	d	What is added where
Sire 1	GS1	--	4/3 to (S1,S1); -2/3 to (GS1,S1); 1/3 to (GS1,GS1)
Sire 2	GS1	Dam 1	2 to (S2,S2); -1 to (GS1,S2), and (D1,S2); 1/2 to (GS1,GS1), (D1,D1), and (GS1,D1)
Sire 3	Sire 1	--	4/3 to (S3,S3); -2/3 to (S1,S3); 1/3 to (S1,S1)
Sire 4	GS2	Dam 1	2 to (S4,S4); -1 to (GS2,S4) and (D1,S4); 1/2 to (GS2,GS2), (D1,D1), and (GS2,D1)
Sire 5	GS2	Dam 1	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)

After all 8 animals are processed the inverse of the relationship matrix is as given below.

	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 then the procedure is somewhat more complicated although Quaas (1976) has developed a rapid method of computing the diagonals of the relationship matrix (which are 1+F) from which the inverse of the whole relationship matrix can be easily computed.

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 relative must be included. Inbreeding is ignored.



If known,

Animal p	Sire s	Maternal grandsire m	Then add what to where[a/]
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)

[a/] Symmetric: for example, if -2/3 to (s,p), then -2/3 to (p,s).

Cochran, W.G. 1951. Improvement by means of selection. Proc. Second Berkeley Symp. Math. Stat. and Prob. 449-470.

Cockerham, C.C. 1954. An extension of the concept of partitioning hereditary variance for analysis of covariance among relatives when epistasis is present. Genetics 39:859.

Comstock, R.E., and H.F. Robinson. 1948. The components of genetic variance in populations. Biometrics 4:254.

Comstock, R.E., H.F. Robinson, and P.H. Harvey. 1949. A breeding procedure designed to make maximum use of both general and specific combining ability. Agron. J. 41:360.

Cruden, Dorothy. 1949. The computation of inbreeding coefficients. J. Heredity 40:248.

Dempster, E.R. and I.M. Lerner. 1950. Heritability of threshold characters. Genetics 35:212.

Dickerson, G.E. 1947. Composition of hog carcasses as influenced by heritable differences in rats and economy of gain. Res. Bull. Iowa Agr. Exp. Sta. 354.

———————— 1962. Implications of genetic-environmental interaction in animal breeding. Animal Prod. 4:47.

———————— 1969. Techniques for research in quantitative genetics. In Techniques and Procedures in Animal Science Research. Am. Soc. An. Sci.

Dickerson et al. 1954. Evaluation of selection in developing inbred lines of swine. Missouri Agr. Expt. Sta. Bull. 551.

Dickerson, G.E. and L.N. Hazel. 1944. Effectiveness of selection on progeny performance as a supplement to earlier culling in livestock. J. Agr. Res. 69:459.

Emik, L.O. and C.E. Terrill. 1949. Systematic procedures for calculating inbreeding coefficients. J. Heredity 40:51.

Everett, R. W. 1975. Income over investment in semen. J. Dairy Sci. 58:1717.

Falconer, D.S. 1952. The problem of environment and selection. Am. Naturalist 86:293.

———————— 1953. Selection for large and small size in mice. J. Genetics 51:470.

Fisher, R.A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. Trans. Royal Soc. Edinburgh 52:399.

Harris, Dewey L. 1970. Breeding for efficiency in livestock production: defining the economic objectives. J. Animal Sci. 30:860.

Hazel, L.N. 1943. The genetic basis for constructing selection indexes. Genetics 28:476.

Hazel, L.N. and J.L. Lush. 1942. The efficiency of three methods of selection. J. Heredity 33:393.

Henderson, C.R. 1952. Specific and general combining ability. In Heterosis, J.W. Gowen (ed.). Iowa State College Press, Ames.

———————— 1953. Estimation of variance and covariance components. Biometrics 9:226.

———————— 1963. Selection index and expected genetic advance. Nat. Res. Council Pub. 982.

———————— 1966. A sire evaluation method which accounts for unknown genetic and environmental trends, herd differences, season, age effects and differential culling. Proc. Symp. on Estimating Breeding Values of Dairy Sires and Cows. U.S.D.A., A.R.S., Beltsville, Md.

———————— 1972. Sire evaluation and genetic trends. Proc. Jay L. Lush Symp. on Animal Breeding and Genetics. Am. Soc. An. Sci. Champaign, Ill.

———————— 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423.

———————— 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 32:69.

Henderson, C.R., H.W. Carter, and J.L. Godfrey. 1954. Use of contemporary herd average in appraising progeny tests of dairy bulls. J. Animal Sci. 13:949.

Henderson, C.R., O. Kempthorne, S.R. Searle, and C.M. von Krosigk. 1959. The estimation of environmental and genetic trends from records subject to culling. Biometrics 15:192.

Kempthorne, O. 1954. The correlations among relatives in a random mating population. Proc. Royal Soc. London B 143:103.

Kempthorne, O. and A.W. Nordskog. 1959. Restricted selection indices. Biometrics 15:10.

Lerner, I.M. and L.N. Hazel. 1947. Population genetics of a poultry flock under artificial selection. Genetics 32:325.

Lin, C. Y. 1978. Review. Index selection for genetic improvement of quantitative characters. Theor. Appl. Genet. 52:49.

Lush, J.L. 1931. The number of daughters necessary to prove a sire. J. Dairy Sci. 14:209.

———— 1933. The bull index problem in the light of modern genetics. J. Dairy Sci. 16:501.

———— 1935. Progeny test and individual performance as indicators of an animal's breeding value. J. Dairy Sci. 18:1.

———— 1940. Intra-sire correlations or regressions of offspring on dam as a method of estimating heritability of characteristics. Proc. Am. Soc. Animal Prod. 293.

———— 1944. The optimum emphasis on dam's records when proving dairy sires. J. Dairy Sci. 27:937.

———— 1947. Family merit and individual merit as bases for selection. Part I. Amer. Nat. 81:241. Part II. Amer. Nat. 81:362.

Mason, I.L. and A. Robertson. 1956. The progeny testing of dairy bulls at different levels of production. J. Agr. Sci. 47:367.

Quaas, R. L. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. Biometrics 32:949.

Quaas, R. L. and L. D. Van Vleck. 1979. Categorical trait sire evaluation by best linear unbiased prediction of future progeny category frequencies. Biometrics

Rendel, J.M. and A. Robertson. 1950. Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. J. Genetics 50:1.

Robertson, A. and J.M. Rendel. 1950. The use of progeny testing with artificial insemination in dairy cattle. J. Genetics 50:21.

———————————— 1954. The performance of heifers got by artificial insemination. J. Agr. Sci. 44:184.

Schaeffer, L.R., L.D. Van Vleck, and J.A. Velasco. 1970. The use of order statistics with selected records. Biometrics 26:854.

Smith, H.F. 1936. A discriminant function for plant selection. Ann Eugen. 7:240.

Tallis, G.M. 1962. A selection index for optimum genotype. Biometrics 18:120.

Van Vleck, L.D. 1967. Effect of artificial insemination on frequency of undesirable recessive genes. J. Dairy Sci. 50:201.

———————— 1969. Relative selection efficiency in retrospect of selected young sires. J. Dairy Sci. 52:768.

———————— 1970. Index selection for direct and maternal genetic components of economic traits. Biometrics 26:477.

_____ 1976. Selection for direct, maternal and grandmaternal components of economic traits. Biometrics 32:173.

_____ 1978. A genetic model involving fetal effects on traits of the dam. Biometrics 34:123.

_____ 1978. Expected economic responses to selection for direct and fetal genetic effects. J. Dairy Sci. 61:1468.

Willham, R.L. 1963. The covariance between relatives for characters composed of components contributed by related individuals. Biometrics 19:18-27.

_____ 1972. Biometrical aspects of maternal effects in animals. J. Animal Sci. 35:1288.

_____ 1972. The role of maternal effects in animal breeding: III. Biometrical aspects of maternal effects in animals. J. Anim. Sci. 35:1288.

Wilton, J.W., D.A. Evans, and L.D. Van Vleck. 1968. Selection indices for quadratic models of total merit. Biometrics 24:937.

Wright, Sewall. 1921. Correlation and causation. J. Agr. Res. 201:557.

_____ 1921. Systems of mating. I through V. Genetics 6:111-178.

_____ 1922. Coefficients of inbreeding and relationship. Amer. Natur. 56:330.

_____ 1934. The method of path coefficients. Ann. Math Stat. 5:161.

Wright, Sewall and H.C. McPhee. 1925. Approximate method of calculating coefficients of inbreeding and relationship from livestock pedigress. J. Agr. Res 31:377.

EXAMINATION QUESTIONS

NOTE WELL: Show all work. Arithmetic errors will not cause you to lose full credit for a problem unless you fail to show how you arrived at the answers. If the correct answer is given but no work is shown, you may lose part credit. Answers may be left in fractional form. You need not find square roots.

First Exam, October 1972

1. A trait has heritability in the narrow sense of .6 and phenotypic variance of 49. What is the expected difference in progeny average between a parent with a phenotypic record of 30 and a parent with a phenotypic record of 20?

2. A researcher has gathered the following pairs of records on a parent's phenotypic record and its progeny phenotypic average.

Pair no.	Frequency of parent phenotypes	Parent's phenotypic record	Progeny average
1	.09	10	7.20
2	.30	8	6.80
3	.37	6	6.34
4	.20	4	5.80
5	.04	2	5.20
		$\mu = 6.4$	$\mu = 6.40$

   a) What is the phenotypic variance of this trait?
   b) What is the additive genetic variance for this trait? (Assume only 2 alleles at one locus are involved.)

3. The regression equation of genetic value on number of B alleles at one locus is $\hat{G} = 4 + 2$(No. of B alleles). What is the additive genetic value for the following genotypes: BB, Bb, and bb?

4. The following table shows the frequency of 3 genotypes at the B locus and the corresponding additive and dominance genetic values.

Genotype	Frequency	Additive genetic value	Dominance generic value
BB	.25	9	-1
Bb	.50	7	+1
bb	.25	5	-1
		$\mu_{G_A} = 7$	$\mu_{G_D} = 0$

   Show numerically that the total genetic variance is equal to the additive genetic variance plus the dominance genetic variance.

5. The following is an arrow diagram of brother-sister mating. Assume that the original parents (a and B) are not related but that A has an inbreeding coefficient of .25 and B is not inbred.

   A C E G

   B D F H

   a) Show all additive relationships among the 8 animals.
   b) What is the inbreeding coefficient of C? Of E? Of G?
   c) Suppose through artificial insemination A is mated to H to produce I. What is the inbreeding coefficient of I? The additive relationship of A and I? The additive relationship of H and I?

6. The following are tables of additive and dominance relationships among animals A, B, C, and D.

Additive relationships

	A	B	C	D
A	$1\frac{1}{4}$	0	$\frac{1}{2}$	$\frac{3}{4}$
B	0	1	$\frac{1}{2}$	$\frac{1}{4}$
C	$\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{1}{2}$
D	$\frac{3}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	1

Dominance relationships

	A	B	C	D
A	1	0	0	$\frac{3}{16}$
B	0	1	0	0
C	0	0	1	$\frac{1}{4}$
D	$\frac{3}{16}$	0	$\frac{1}{4}$	1

   Given: $\sigma^2_{10} = 64$, $\sigma^2_{20} = 16$, $\sigma^2_{01} = 48$, $\sigma^2_{11} = 8$, and $\sigma^2_E = 64$.
   a) What is heritability in the narrow sense? In the broad sense?
   b) What is the genetic covariance between A and D? Between C and D? Between B and C? Between A and A? Between B and B?

Second Exam, November 1972

1. The heritability of a trait is .3, repeatability is .4, and phenotypic variance is 100. Set up equations (numerically, but do not solve) to find the weights to evaluate an animal that has records on the following relatives:

   X_1, average of two records on the dam = 400 ;
   X_2, average of six records on the maternal granddam = 200 ;
   X_3, average of two records on each of 20 paternal half-sibs = -300 ;
   X_4, average of single records on each of 10 full sibs = 500.

2. The additive genetic value for the following 6 animals has been predicted by use of selection indexes. The r_{TI} for each index is also given. $h^2 = .4$, $r = .6$, and variance of single records is 400.

Animal	Index Value	r_{TI}
A	40	.70
B	50	.60
C	10	.50
D	-60	.40
E	-40	.80
F	30	.60

   a) Which two animals should be selected to give the greatest superiority in predicted additive genetic value?
   b) What is the expected superiority in additive genetic value if those two are selected out of the six?

3. The following additive relationship table is known.

	A	B	C	D
A	1	1/4	0	0
B	1/4	$1\frac{1}{4}$	0	0
C	0	0	1	1/4
D	0	0	1/4	1

The following records are available for the four animals.

Animal	No. of Records	Ave. of Records
A	2	3
B	1	4
C	1	-5
D	3	2

$h^2 = .25$, $r = .50$, and variance of single records = 36. Predict the additive genetic value of animals B and D.

4. The following information is available to predict the additive genetic value of a sire A.

X_1 = 100 = a single record of a progeny with sire A and dam D
X_2 = 200 = a single record of another progeny also with sire A and dam D
X_3 = -300 = a single record of D

D is not related to A. $h^2 = .4$, $r = .8$, and variance of single records = $(100)^2$.

a) Predict the additive genetic value for sire A.

b) What is the r_{TI} for this index? Note r_{TI}.

c) What would be the r_{TI} if the record of D is not considered?

5. A researcher has determined that the environmental covariance between full sibs is $(.2)\sigma_X^2$ where $\sigma_X^2 = 400$ is the phenotypic variance for the trait; heritability is .60. A breeder wants to predict the additive genetic value of an animal with the following information.

X_1 = 10, the average of single records of five full sibs of the animal being evaluated.
X_2 = 20, a single record of the animal.

a) Try to give the breeder what he wants.
b) How would you predict the additive genetic value of another of the full sibs if its records was -10?

6. Assume $X_1 = 30$ is the average of single records of 10 progeny (paternal half-sibs) of sire A, and $X_2 = 60$ is the average of single records of 5 progeny (paternal half-sibs) of sire B. The dams are all unrelated as are sires A and B. $h^2 = .4$, $r = .5$, and variance of single records = 2500.

a) Find the weights for the index, $I = b_1X_1 + b_2X_2$ to predict the difference in additive genetic value between sires A and B, i.e., predict $T = T_A - T_B$.
b) Predict the difference in additive genetic value between sires A and B.
c) Calculate the r_{TI}^2 for this index.
d) Calculate the average squared prediction error for this index, i.e., $E[(I-T)^2] = V[I - (T_A - T_B)]$.

Final Exam, December 1972

Some of the queations are rather lengthy and represent lectures more than difficult problems so bear with them and do not give up. Questions 1, 2, and 3 are relatively easy. Part c of #4, part d of #5, part b of #6, and half of part c of #6 are a little more difficult but do not account for many points.

1. The heritability of a trait is .3, repeatability is .4, and phenotypic variance is 100.

a) Set up equations (numerically, but do not solve) to find the weights to predict additive genetic value for evaluating animals thar have records on the following relatives.

X_1, average of <u>three</u> recorda on the sire = 400 ;
X_2, average of <u>four</u> records on the paternal granddam = 200 ;
X_3, average of <u>single</u> records on each of 5 maternal half-sibs = 200 ;
X_4, average of <u>two</u> records on each of 8 full sibs = 500.

b) Can you compare predicted additive genetic values (indexes) of such animals with those for animals that have only X_1 and X_2? In words, what is the procedure for evaluating animals that have only X_1 and X_2 if they are to be compare to animals having X_1, X_2, X_3, and X_4?

2. Additive genetic value for several unrelated males and females has been predicted from records of relatives by the selection index procedure. $h^2 = .2$, $r = .6$, and $\sigma_X^2 = (2000)^2$.

Male	Index	r_{TI}	Female	Index	r_{TI}
A	500	.90	P	100	.60
B	1000	.70	Q	-200	.50
C	-500	.80	R	500	.50
			S	300	.60
			T	-400	.50

a) What is the predicted difference in average additive genetic value between progeny of males A and C if each sire were mated to all five females, i.e., A with P, Q, R, S, T; and C with P, Q, R, S, T?
b) What is the predicted additive genetic value of a progeny obtained from mating male B to female R? B to female Q?

3. Two inbred lines have been developed so that all animals in line 1 have inbreeding coefficient $F_1 = .6$ and all animals in line 2 have inbreeding coefficient $F_2 = .4$. The lines have reached a constant inbreeding coefficient so that all animals within a line are related to each other by 2F so that mating of random males and females within a line results in a progeny with inbreeding coefficient (1/2)(2F) = F. Assume that the two lines are unrelated and are to be crossed as shown.

A and B are random linecross progeny of the cross between lines 1 and 2 since we assume that each parent has only one progeny.

a) Complete the additive relationship table for $F_1 = .6$ and $F_2 = .4$ where ℓ_1 and ℓ_1' are animals from line 1 and ℓ_2 and ℓ_2' are animals from line 2.

	ℓ_1	ℓ_2	ℓ_1'	ℓ_2'	ℓ_1, ℓ_2 A	ℓ_1', ℓ_2' B
ℓ_1	$1+F_1=1.6$	0	$2F_1=1.2$	0		
ℓ_2	0	$1+F_2=1.4$	0	$2F_2=.8$		
ℓ_1'	$2F_1=1.2$	0	$1+F_1=1.6$	0		
ℓ_2'	0	$2F_2=.8$	0	$1+F_2=1.4$		
A						
B						

b) What is the additive relationship between linecross animals, e.g., A and B?

c) What is the dominance relationship between A and B?

d) Assume $\sigma^2_{10} = 100$, $\sigma^2_{01} = 50$, $\sigma^2_{11} = 25$, $\sigma^2_{20} = 60$, and $\sigma^2_E = 265$. What is the genetic covariance between random linecross animals, i.e., between A and B? If you are not sure of the answer for parts b and c use $a_{AB} = .25$ and $d_{AB} = .12$.

4. Often the purpose of measuring traits early in life is said to be that they will predict performance for an economically important trait which cannot be measured until later. If such prediction is reasonably accurate, then the reduced generation interval will result in more progress for the economic trait than will selection based on a record of the economic trait. An example is mastitis in dairy cattle which becomes a major problem in third and later lactations.

Assume X_1 and X_2 are measure of traits 1 and 2 early in life and X_3 is a measure of the economic trait. Compare genetic progress per year genetically in improving trait 3 from two plans:

a) (1) Select for additive genetic value of trait 3 based on X_1 and X_2. The top 70 of 100 will be selected and the generation interval is 4 years.

b) (2) Select for trait 3 using only X_3. Since many of the animals will have died by then, the top 70 of 90 will be selected. The generation interval is 8 years.

c) Also, what are the expected correlated genetic responses in trait 1 for plans (1) and (2).

The following variance and covariances are known.

Phenotypic: $\sigma^2_{X_1} = 64$, $\sigma_{X_1X_2} = 12$, $\sigma_{X_1X_3} = -13$, $\sigma^2_{X_2} = 144$, $\sigma_{X_2X_3} = 0$, and $\sigma^2_{X_3} = 196$.

Genetic: $\sigma^2_{G_1} = 40$, $\sigma_{G_1G_2} = -6$, $\sigma_{G_1G_3} = 18$, $\sigma^2_{G_2} = 36$, $\sigma_{G_2G_3} = 36$, and $\sigma^2_{G_3} = 49$.

5. Additive genetic value has been predicted for a male from his progeny average (paternal half-sibs), X_1, and for a female from her record, X_2, and her dam's record, X_3. $h^2 = .4$, $r = .6$, and the variance of single records is $\sigma^2_X = (1000)$, $\sigma^2_{10} = 400{,}000$, $\sigma^2_{10} = 630$. The index, index values, and r^2_{TI} are given.

Male: $I = (5/8)X_1 = 810$; $r^2_{TI} = 5/16$

Female: $I = (3/8)X_2 + (1/8)X_3 = 390$; $r^2_{TI} = 7/16$

a) What is the predicted additive genetic value for an unborn progeny of mating this unrelated male and female?

b) The r_{TI} for an unborn progeny of mating this unrelated male and female is $(1/2)\sqrt{r^2_{TI}}$ for sire + r^2_{TI} for dam. What is the probability that the additive genetic value of the progeny will exceed -100? If you are not sure of the answer to patt a use $I_{progeny} = 400$.

c) What is the probability that the record of the progeny will be less than -1000? If you are not sure of the answer to part a, use $I_{progeny} = 400$.

d) Show either numerically for this example or indicate algebraically why the r^2_{TI} for a progeny with no records equals the sum of the r^2_{TI} for the sire and the r^2_{TI} for the dam divided by four when the sire and dam are unrelated, i.e., $r^2_{TI_{progeny}} = (r^2_{TI_{sire}} + r^2_{TI_{dam}})/4$. Some of the following squsre roots not included can be approximated from these.

6. Prediction error can be defined as $e_\alpha = T_\alpha - I_\alpha$ where T_α is true value and I_α is the index estimate of true value for animal .

Assume: $\sigma^2_{X_1} = 100$, $\sigma^2_{X_2} = 100$, $\sigma_{X_1X_2} = 20$, $\sigma_{X_1T} = 40$, $\sigma_{X_2T} = 20$, and $\sigma^2_T = 40$.

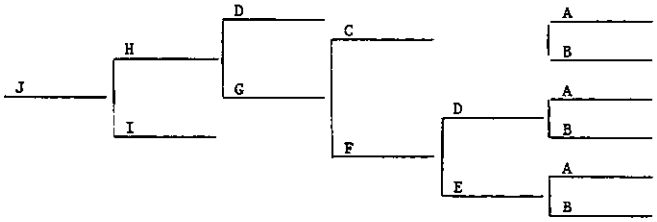
a) The index which minimizes squared prediction error is $I = (3/8)X_1 + (1/8)X_2$. What is the average squared prediction effor for this best index?

b) What is the variance of prediction error for the index $I = 6X_1 + 2X_2$ which is obviously not the index that comes from solving the selection index equations? Compare the answers for parts a and b. Which is larger? A short answer, why?

c) Compare r^2_{TI} for the indexes $I = 6X_1 + 2X_2$ and $I = (3/8)X_1 + (1/8)X_2$. Why are they the same or why are they different? Note r^2_{TI} rather than r_{TI} to make the computations easier.

First Exam, October 1973

1. A dairyman has had some trouble computing the inbreeding to be expected from some possible matings. Please compute the additive relationships among the following animals. Also indicate to him which animals are inbred and their inbreeding coefficients. Note well: A and B are full brother and sister. C and D are identical twin bulls.



2. With the help of a miracle, the following genotypes for ten loci have been determined for two animals. The genes are identified by their origin.

a) What is the estimate of inbreeding for animal X?
b) For animal Y?
c) What is the additive relationship between X and Y?
d) What is the dominance relationship between X and Y?

Locus	Animal X	Animal Y
A	a_1a_2	a_3a_6
B	b_7b_7	b_7b_8
C	c_2c_3	c_1c_3
D	d_4d_4	d_4d_8
E	e_6e_9	e_6e_9
F	$f_{11}f_{12}$	f_1f_2
G	g_8g_6	g_6g_9
H	h_1h_2	h_3h_4
I	i_1i_6	i_1i_8
J	$j_{10}j_{12}$	$j_{10}j_{12}$

3. To set up the selection index equations, the genetic covariances between relatives are often deeded. Assume $\sigma^2_{10} = 32$, $\sigma^2_{20} = 16$, $\sigma^2_{01} = 8$, $\sigma^2_{11} = 24$, and $\sigma^2_E = 120$. The additive relationships and dominance relationships among a group of beef animals are given in the tables below.

Additive Relationships

	A	B	C	D	E
A	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$
B	$\frac{1}{2}$	$1\frac{1}{4}$	$\frac{5}{8}$	$\frac{1}{4}$	$\frac{1}{8}$
C	$\frac{1}{2}$	$\frac{5}{8}$	$1\frac{1}{4}$	0	0
D	$\frac{1}{4}$	$\frac{1}{4}$	0	1	$\frac{1}{4}$
E	$\frac{1}{8}$	$\frac{1}{8}$	0	$\frac{1}{4}$	1

Dominance Relationships

	A	B	C	D	E
A	—	$\frac{1}{4}$	0	0	0
B	$\frac{1}{4}$	—	0	0	$\frac{1}{32}$
C	0	0	—	0	0
D	0	0	0	0	$\frac{1}{16}$
E	0	$\frac{1}{32}$	0	$\frac{1}{16}$	—

a) What is the genetic covariance between a record of A and a record of D?
b) Between A and B?
c) Between C and D?
d) Suppose B and E were mated. What would be the inbreeding of the offspring?

4. The model for a record on some trait is $P_i = G_i + E_i$ where $\mu_G = \mu_E = 0$; $\sigma^2_G = \sigma^2_{10} + \sigma^2_{20} + \sigma^2_{01} = 16 + 8 + 4$; $\sigma^2_E = 72$; and $\sigma_{GE} = 0$.

a) What is the variance of single recodds, i.e., σ^2_P?
b) What is heritability for this trait?
c) Suppose $\sigma_{GE} = -16$. What would be the variance of single records?
d) Suppose $\mu_G = 120$ and $\mu_E = 0$. What is the varisnce of single records?
e) Suppose $\mu_G = 40$ and $\mu_E = 110$. What is the variance of single records?

5. The model for a record on some trait is $P_i = G_i + E_i$ where $\mu_G = \mu_E = 0$; $\sigma^2_G = \sigma^2_{10} + \sigma^2_{20} + \sigma^2_{01} = 16 + 8 + 4$; $\sigma^2_E = 72$; and $\sigma_{GE} = 0$.

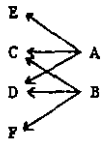
A horse breeder wants to select on an index which is the average of the parents' single records plus the animal's own record, i.e., $I = [(P_1+P_2)/3] + P_3$, where P_1 is the record of the sire, P_2 is the record of the dam, and P_3 is the record of the animal.

a) What is the variance of this index? Assume the sire and dam are unrelated and are not inbred.

b) What is the covariance between the index and the genetic value of the animal?

Second Exam, November 1973

1. Single records are available on animals A, B, D, E, and F. They are related to animal C as indicated in the arrow pedigree shown below. The records are: $X_A = 30$ $X_B = -10$ $X_D = 40$ $X_E = 10$ $X_F = -5$.



Heritability of the trait is .4, repeatability is .6, and the phenotypic variance is 500.

a) Indicate clearly and numerically as much as possible, but <u>without solving any equations</u>, how to predict the additive genetic value of animal C.

b) Repeat a), but show how to predict the additive genetic value of animal F.

c) What assumptions are necessary so that your procedure will minimize average squared prediction error?

2. A dairyman wants to predict <u>twice</u> the additive genetic value for some type trait for several bull calves from the type trait scores of each calf's four unrelated grandparents. Let X_{PGS}, X_{PGD}, X_{MGS}, and X_{MGD} be the records of the four grandparents. The best index is

$$I = .18X_{PGS} + .18X_{PGD} + .18X_{MGS} + .18X_{MGD} .$$

Assume heritability = .36, repeatability = .49, and variance of single records = 9.

a) What is the r_{TI} for this index?

b) What is the average squared prediction error for this index?

Suppose six of the bull calves have grandparents with records as shown:

Calf	X_{PGS}	X_{PGD}	X_{MGS}	X_{MGD}
A	3	0	-2	3
B	0	2	1	0
C	-2	-1	0	3
D	-1	6	-4	-1
E	1	2	0	2
F	-1	3	-2	-2

c) Rank the six calves for twice their additive genetic value for the type trait.

d) Suppose you choose what you think are the best two calves for this trait. What is the expected superiority in twice the additive genetic value for the selected two above the initial group of six?

DO EITHER QUESTION 3 OR QUESTION 4.

3. A swine breeder wants you to predict the average of single records of 10 unborn progeny from mating sire, S, and dam, D. Five previous full sib progeny of S and D have averaged +5. Assume heritability = .10, repeatability = .15, the environmental covariance between the full sibs already born is $.2\sigma_X^2$, the environmental covariance between the unborn progeny will be $.2\sigma_X^2$, and the variance of single records, $\sigma_X^2 = .36$. Assume no environmental covariance between the 5 full sib progeny and the unborn progeny.

a) Predict the average of 10 future progeny.

b) What is the σ_T^2 for this procedure?

DO EITHER QUESTION 3 OR QUESTION 4.

4. Suppose each animal to be evaluated has two records and that the average of twenty full sibs each with two records is also known. You are to predict a future record of each of the following four animals.

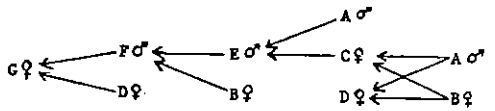
Animal	Ave. of animal's two records	Ave. of 2 records on each of 20 full sibs of the animal
A	4	3
B	2	4
C	-6	10
D	8	-2

Heritability of the trait is .25, repeatability is .50, the environmental covariance between full sibs is $.20\sigma_X^2$, and the variance of single records is $\sigma_X^2 = 100$.

5. A bull has sired a number of daughters resulting in a rather high index value for his additive genetic value. The breed organization, however, wants to present a Breeder of Merit award to the bull's owner: but they want to be 90% sure that if the bull sires an infinite number of additional progeny that his additive genetic value will not fall below +600 pounds. The additive genetic variance for the trait is $(1200 \text{ lb.})^2$, and the phenotypic variance is $(2400 \text{ lb.})^2$. Suppose the r_{TI} is .8; how high should the index value be before the award is presented?

Final Exam, December 1973

1. a) Compute the additive relationships among the animals shown in the arrow diagram. Indicate which are inbred and what their inbreeding coefficients are.



b) Assume that for some trait the direct additive genetic variance $\sigma^2_{G_D}$ = 80; the maternal additive genetic variance, $\sigma^2_{G_M}$ = 40; and covariance between the additive direct and additive maternal genetic effects $\sigma_{G_D G_M}$ = -50. Also assume the variance of single record σ^2_P = 200.

i) What is the environmental variance for this trait?
ii) What is the covariance between a record of D and a record of E?
iii) What is the covariance between a record of C and a record of F?
iv) What is the covariance between a record of G and a record of F?

2. A breeder is selecting for a defined true value which has variance σ^2_T = 52.5. There are available three sources of information, X_1, X_2, and X_3. The equations to find the selection index weights are: $100b_1 + 50b_2 + 25b_3 = 52.5$; $50b_1 + 100b_2 + 50b_3 = 45$; and $25b_1 + 50b_2 + 100b_3 = 30$. The weights are: $b_1 = .4$, $b_2 = .2$, and $b_3 = .1$. Five animals have the following information.

Animal	X_1	X_2	X_3
A	5	-4	3
B	3	4	-1
C	10	-5	-5
D	-5	4	4
E	2	10	1

a) Rank the five animals for predicted true value.
b) What is the expected superiority in true value of the three you think are best over the two you think are poorest?
c) What is the average squared prediction error for this index?
d) Suppose another animal, F, has an index value of 3.0. What is the probability that the true value of animal F is between 0 and 4?

3. A sheepman has the following information for one trait available on all of his sheep. He wants you to develop an index to rank his animals for a future record. X_1 = animal's own record, X_2 = average of two record of the dam, X_3 = average of single records of two maternal half-sisters of the animal, and X_4 = single record of a full sib of the animal. Variance of single records = 100; heritability = .40, and repeatability = .60. Environmental covariance between a record of a mother and her progeny = 20% of the variance of single records, and environmental covariance between records of full sibs = 25% of the variance of single records.

Show clearly how to develop the desired index. Do not solve any equations.

4. A breeder has decided that the economic value of trait 1 is \$100 per standard deviation and of trait 2 is \$49 per standard deviation. Heritabilities of the traits are h^2_1 = .25 and h^2_2 = .50. He decides to use the index for economic additive genetic value I = (\$10)(.25)X_1 + (\$7)(.5)X_2. The phenotypic and genetic variance and covariances are: $\sigma^2_{X_1}$ = 100, $\sigma^2_{X_2}$ = 49, $\sigma_{X_1X_2}$ = 35, $\sigma^2_{G_1}$ = 25, $\sigma^2_{G_2}$ = 24.5, and $\sigma_{G_1G_2}$ = 15.

a) What is the r_{TI} for the index?
b) What is the expected superiority in economic value for selection of the top 2 out of 15 animals?

5. The government of a country with a shortage of fat and protein wants to select dairy cows to improve yield of fat and yield of protein. They have decided that a pound of protein is worth \$6.00 and that a pound of fat is worth \$2.00 in their currency. The following variances and covariances are known.

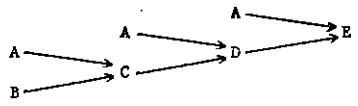
	Phenotypic	Genetic	Permanent Environmental	Temporary Environmental
Protein-fat covariance	6400	1800	1400	3200
Protein variance	6400	1600	1600	3200
Fat variance	10,000	3600	2500	3900

a) As their consultant you are asked to prepare an index to predict a future economic record for cows that have a first record for both fat and protein; that is, X_1 = first protein record and X_2 = first fat record. Find the appropriate weights.

b) Describe clearly how to determine the index to predict an economic record for a daughter of cow with X_1 and X_2. Do not solve the equations.

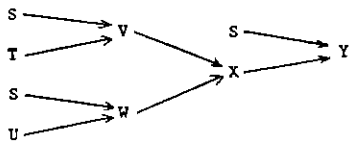
First Exam, October 1974

1. a) One horse breeder has followed a close linebreeding (inbreeding) program by continued mating to sire A.



Show all additive relationships among the five horses and give the inbreeding coefficient for those that are inbred.

b) Another breeder follows a less intense linebreeding program to sire S.



Show all additive relationships among the severn horses and give the inbreeding coefficient for those that are inbred.

c) Another breeder likes the results both (a) and (b) have achieved. He buys E and Y. He asks you what the inbreeding coefficient will be for progeny resulting from mating E to Y. Your answer is?

2. Plant breeders have an advantage in developing inbred lines because of the large number of plants that can be used, the shorter generation interval, and the ability of selfing (same plant provides the male and female gametes for fertilization). The closest analogy in animal breeding to selfing would be if identical twins could be mated. Since identical twins cannot be mated, the fastest practical method of inbreeding in animals is to make continuous full brother-full sister (full sib) matings. In rhe following problems, assume the base population is noninbred and unrelated. Only half credit will be given for the correct answer unless all work is shown.

a) Calculate the inbreeding coefficient after 3 generations of full sib matings.

b) Calculate the inbreeding coefficient after 3 generations of selfing.

3. The following covariances between 3 sets of relatives have been computed for a trait: between parent and progeny, 56; between paternal half-sibs, 24; between full sibs, 64. From these values estimate:

a) σ^2_{10}, σ^2_{01}, σ^2_{20}. b) σ^2_{10}, σ^2_{20}, σ^2_{11}.

Information for questions 4 and 5

The variances of averages are often needed to set up the equations to solve for the selection index weights. For some trait, assume the following: phenotypic variance = 400; additive genetic variance = 100; dominance genetic variance = 60; additive by additive genetic variance = 40; environmental covariance between parent and progeny = 50; environmental covariance between full sibs = 40; environmental covariance between paternal half-sibs = 30; permanent environmental variance = 50; temporary environmental variance = 150; X_1 is the average of single records of 10 full sib progeny of sire A and dam B; X_2 is the average of single records of 5 paternal half-sib progeny of sire A (all have different dams, none of which is B); and $X_3 = (10X_1+5X_2)/15$, i.e., the average of the 15 progeny of sire A.

4. a) Compute the variance of X_1.

b) Compute the variance of X_2.

c) Compute the covariance between X_1 and X_2.

d) Compute the variance of X_3. (Note this kind of average is not used in the selection index since all records do not have the same covariance with true value.)

5. a) What is heritability for this trait?
b) What is repeatability for this trait?
c) What is the covariance between parent and progeny records?
d) What is the covariance between X_2 and the genetic value of A?

Second Exam, November 1974

1. A and B are maternal half-sibs. Two records of A average +12. Four records of B average +20. Assume $h^2 = .4$, $t = .6$, and phenotypic variance = $(30)^2$.

a) Predict the difference in additive genetic value of A and B, i.e., $G_A - G_B$.

b) Predict a future record of A and a future record of B.

2. For some trait, assume that: variance of single records = 100; environmental covariance between identical twins = 30; environmental covariance between full sibs = 10; environmental covariance between half-sibs = 5; additive genetic variance = 16; dominance genetic variance = 8; additive by dominance genetic variance = 4; permanent environmental variance = 32; and temporary environmental variance = 40. You are to predict the genetic value of animal 1 from: X_1 = record on the animal; X_2 = records on the animal's identical twin; X_3 = record on a full sib: and X_4 = average of 2 records on each of 20 paternal half-sibs of the animal. Set up, but do not solve, the equations to find the weights for X_1, X_2, X_3, and X_4. You must show the coefficients both symbolically and numerically.

3. A breeder has available to use in selection a record, X_1, on each animal to be selected and a record, X_2, on the dam of each animal. You know that $\sigma^2_{X_1} = 25$, $\sigma^2_{X_2} = 25$, $\sigma_{X_1X_2} = 5$, $\sigma_{X_1T} = 10$, $\sigma_{X_2T} = 5$, and $\sigma^2_T = 20$. This would correspond to T = additive genetic value with $h^2 = .4$. You tell him that the best index is $I = (3/8)X_1 + (1/8)X_2$. He tells you that he has been using the index $I = (75X_1+25X_2)/100$.

a) What is the r^2_{TI} for the best index? For his index?

b) What is average of squared prediction errors for the best index? For his index?

c) What is the variance of the best index? Of his index?

d) What is the correlation between the best index and his index?

4. Sire A has an index of +200 lb for additive genetic value for milk yield. The r^2_{TI} is .84. Sire B has an index of +1200 lb for additive genetic value for milk yield. The r^2_{TI} is .64. Assume $h^2 = 1/4$, $r = 1/2$, phenotypic variance = $(2000)^2$, and that A and B are unrelated as are their daughters. Their index values are based on records of daughters that are related only through the sire.

a) What is the probability that the additive genetic value of A exceeds 1000 lb?

b) What is the probability that a record of a future daughter of A will exceed +2500 lb?

c) What is the probability that the additive genetic value of A exceeds that of B?

d) What fraction of the records of future daughters of A would be expected to be higher than the lowest half of future daughters of B?

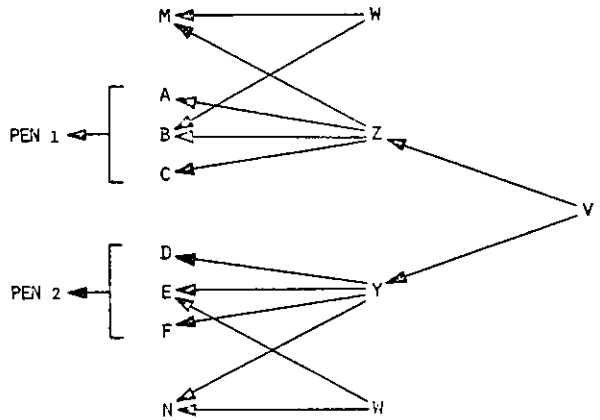
5. Compare the following dairy breeding programs for yearly progress in additive genetic value for milk yield. Assume $h^2 = 1/4$, $r = 1/2$, and phenotypic variance = (2500 lb) .

   a) The careful program? The AI stud will not use sires to sire sons until the records of a second (large) group of daughters is available. Assume $r_{TI} = 1$. The generation interval for sires of sires is lengthened by 3 1/2 yr to 10 yr. Assume the best 5 of 75 are selected. They also will not select cows to produce new sons unless that have at least 5 records. Assume selection is based on her own 5 records. The generation interval for dams of sires becomes about 8 1/2 yr. Assume selection of the top 5%. Assume sires of replacement heifers are selected on the basis of 35 daughter records and that only the best 5 out of 25 bulls are selected. The generation interval is 7 yr. Dams of replacement heifers are selected on the basis of 2 records out of the top 90%. Assume a generation interval of 6 yr.

   b) The dangerous program? The AI stud selects sires of sires on the basis of single records of the first group of 35 daughters. The generation interval will be 6 1/2 yr. More bulls are alive to select from so assume the best 5 of 100 are selected to be sires of sires. Dams of young sires will be selected on an average of only 2 of her records. Assume selection is from the top 3%. The generation interval will be 6 yr. Sires of dams and dams of dams will be selected as for the careful program (a).

Final Exam, December 1974

1. Pedigree for this problem:



Assume heritability is .40, repeatability is .60, and the phenotypic variance is 2500. You want to buy 3 animals from a breeder. He will sell you either Pen 1 containing A, B, and C or Pen 2 containing D, E, and F. You want to buy the pen having the greater average additive genetic value. A records, $X_B = 50$, on animal B and a record, $X_N = 100$, on animal N are available.

   a) Predict the difference in average additive genetic value between Pen 1 and Pen 2.

   b) What is the variance of what is to be predicted?

   c) What is the variance of prediction error?

The following is information for questions 2 and 3:
Assume heritability = .25, repeatability = .40, variance of single records = $(2000 \text{ lb})^2$. Assume also that animals A, B, C, D, and E are unrelated.

Animal	Records Available
A	Record on A = +3000; record on A's dam = -500
B	Average of single records of 20 paternal half-sibs of B = +400
C	Record on dam of C = +2000
D	Record on dam of D = -1000; average of single records of 20 paternal half-sibs of D = +500
E	Two record on dam of E average +500; record on a maternal half-sib of E = -4000
F	Prediction of future record = +500; r_{TI} = .30
G	Prediction of future record = +100; r_{TI} = .30
H	Prediction of future record = -400; r_{TI} = .30
I	Prediction of future record = +1000; r_{TI} = .30
J	Prediction of future record = =300; r_{TI} = .30

2. a) Predict a future record for each of animals A, B, C, D, and E.
   b) What is the probability that a future record of C will be less than +4000?
   c) What is the probability that a future record of A will exceed a future record of C by 3000 or more?

3. a) Suppose out of animals F, G, H, I, and J that 2 animals are to be selected. What would you predict to be the average superiority in a future record of the selected 2 over the average of the group of 5?

   b) Another breeder is going to use a selection index to predict future records. The r_{TI} of the index is .30. He will select the top 40% based on that index. What is the expected superiority in future records of the selected group?

4. Falconer has pointed out that the expressions of a triat in 2 greatly different environments can be considered to be 2 separate traits that have a genetic correlation. Suppose that the 2 environments are: 1) the temperate US and 2) the tropical area of Puerto Rico, and that we are interested in improving milk production. Assume: $h_1^2 = .25$; $h_2^2 = .20$; $\sigma_{P_1}^2 = (2000 \text{ lb})^2$; $\sigma_{P_2}^2 = (1000 \text{ lb})^2$; $r_{G12} = .90$. Compare the following 2 breeding plans for Puerto Rico. Assume the generation interval is the same for both.

a) Bulls are tested in the US on the basis of the average of single records of 50 paternal half-sister daughters. Each generation the best 5 bulls out of 50 tested in the US are selected to be used in Puerto Rico.

b) Bulls are tested in Puerto Rico to be used in Puerto Rico on the basis of the average of single records of 25 paternal half-sister daughters. Each generation the best 5 bulls out of 15 tested in Puerto Rico are selected to be used in Puerto Rico.

5. For weaning weight of beef calves, the following information is given: σ_P^2 = variance of single records = (40 lb)^2; $\sigma_{G_D}^2$ = additive genetic variance for direct effects = $.30\sigma_P^2$; $\sigma_{G_M}^2$ = additive genetic variance for maternal effecrs = $.40\sigma_P^2$; $\sigma_{G_D G_M}$ = additive genetic covariance between direct and maternal effects = $-.30\sigma_P^2$.

a) Find the weighting factors for predicting the additive genetic value for direct effects from X_1, a bull calf's own weaning weight, and X_2, the average weaning weight of 35 of his paternal half-sibs.

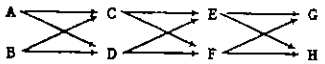
b) Find the weighting factors for predicting the additive genetic value for maternal effects from X_3, a heifer calf's own weaning weight, and X_4, the weaning weight of her dam.

c) Assume bull calves are selected on the basis of (a) from the top 5 of of 50 and heifer calves are selected on the basis of (b) from the top 30 of 50. What would be the expected progeny average of mating the top bulls to the top heifers?

First Exam, October 1975

1. Wednesday

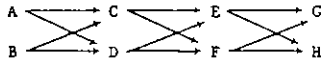
A breeder of dogs wants to develop an inbred line of dogs in spite of all your protests about the dangers of inbreeding. He will practice brother-sister (full sib) mating as described below. A and B are full sibs (brother and sister) from mating of noninbred parents. The inbreeding coefficient of A is $F_A = 1/4$. It is easier to express all calculations as fractions rather than as decimal numbers.



a) Show additive relationships among the eight animals.
b) What will be the inbreeding coefficient of G?
c) What will be the inbreeding coefficient of progeny obtained from mating G and H?

Thursday

A breeder of cats wants to develop an inbred line of cats in spite of all your protests about the dangers of inbreeding. He will practice brother-sister (full sib) mating as described below. A and B are full sibs (brother and sister) from mating of noninbred parents. The inbreeding coefficient of B is $F_B = 1/4$. It is easier to express all calculations as fractions rather than as decimal numbers.



a) Show additive relationships among the eight animals.
b) What will be the inbreeding coefficient of G?
c) What will be the inbreeding coefficient of progeny obtained from mating G and H?

2. A goat breeder believes that additive genetic effects, dominance genetic effects, and additive by additive genetic effects are important for some trait. The variances of these effects are assumed to be:

additive genetic variance = 64,
dominance genetic variance = 32, and
additive by additive genetic variance = 16; also,
variance of environmental effects = 88.

There are also no environmental covariances among records of relatives. Covariances among relatives are needed in order to find the best index for selection. Assume the following dominance and additive relationships among relatives 1, 2, 3, and 4.

Additive Relationships

	1	2	3	4
1	-	1/2	0	1/2
2	1/2	-	1/8	1/4
3	0	1/8	-	1/4
4	1/2	1/4	1/4	-

Dominance Relationships

	1	2	3	4
1	-	1/4	0	0
2	1/4	-	1/32	0
3	0	1/32	-	0
4	0	0	0	-

If X_1 is the average of 2 records on relative 1, X_2 is a record on relative 2, X_3 is a record on relative 3, and X_4 is the average of 3 records on relative 4,

a) What is heritability for this trait?
b) What will be the appropriate covariances for the selection index equations?

$\sigma_{X_1X_2}$ = $\sigma_{X_1X_3}$ = $\sigma_{X_1X_4}$ =

$\sigma_{X_2X_3}$ = $\sigma_{X_2X_4}$ = $\sigma_{X_3X_4}$ =

3. A swine breeder suspects that there is a strong environmental covariance for some trait among full sibs (full brothers and sisters) in the same litter. The breeder is willing to assume that the only genetic effects are additive genetic effects and that there is no environmental covariance among pigs with the same father but with different mothers (paternal half-sibs) which are born at the same time of year.

A statistician has gathered some data and has calculated the covariance between full sibs and the covariance between paternal half-sibs:

full sib covariance = 30,
paternal half-sib covariance = 10, and
variance of single records = 100.

a) Estimate heritability.
b) Estimate the environmental covariance between full sibs.

Another breeder, however, believes that dominance effects are also important.

c) Can dominance genetic variance be estimated from the data given above? If yes, what is the estimate? If no, what can be estimated?

4. Wednesday

The following animals all have relatives of type A and B with a single record each, X_A and X_B.

	Record on Relative	
Animal	A	B
1	+20	-10
2	+40	-30
3	+8	+20
4	-8	+50

The following information is known:

$\sigma_X^2 = 100$; $\text{Cov}(X_A, X_B) = 20$;

$\text{Cov}(X_A, T) = 22$; $\text{Cov}(X_B, T) = 14$; $\sigma_T^2 = 50$.

$I = .2X_A + .1X_B$.

a) Rank the animals according to the index.
b) Will your rank be the same as the rank for true value? Why or why not?
c) What is the variance of this index?
d) What is the variance of prediction errors, $I-T$, for this index?

Another breeder decides the selection emphasis in tha above index is wrong and uses the index $I^* = X_A + 2X_B$.

e) What is the variance of his index?
f) What is the variance of prediction errors, I^*-T, for his index?

Thursday

The following animals all have relatives of type A and B with a single record each, X_A and X_B.

	Record on Relative	
Animal	A	B
1	+20	-10
2	+40	-30
3	+8	+20
4	-8	+50

The following information is known:

$\sigma_X^2 = 100$; $\text{Cov}(X_A, X_B) = 20$;

$\text{Cov}(X_A, T) = 22$; $\text{Cov}(X_B, T) = 14$; $\sigma_T^2 = 50$.

$I = .2X_A + .1X_B$.

a) Rank the animals according to the index.
b) Will your rank be the same as the rank for true value? Why or why not?
c) What is the variance of this index?
d) What is the correlation between this index, I, and true value, T?

Another breeder decides the selection emphasis in the above index is wrong and uses the index $I^* = X_A + 2X_B$.

e) What is the variance of his index?
f) What is the correlation between his index, I^*, and true value, T?

Second Exam, November 1975

1. A dairyman friend knows that you are now an expert in determining genetic merit of animals from their records and records of their relatives. He wants you to buy six cows for him that have highest possible genetic value for milk yield. The only cows available and information on their evaluations are given below. The owner will sell you two groups, but you cannot make any substitutions. All groups are equally priced. Heritability for milk is .25, repeatability is .50, and the phenotypic variance is $(2000 \text{ lb})^2$.

	Cow	Selection Index for Additive Genetic Value for Milk Yield (lb)	Average Index for Group (lb)	r_{TI}
	A1	+1000		.65
Group A	A2	+500	(400)	.60
	A3	-300		.40
	B1	+1200		.70
Group B	B2	0	(300)	.20
	B3	-300		.35
	C1	+2000		.50
Group C	C2	+500	(500)	.60
	C3	-1000		.25
	D1	+450		.60
Group D	D2	+450	(450)	.50
	D3	+450		.40
	E1	+3000		.70
Group E	E2	-1200	(100)	.70
	E3	-1500		.40

a) Which two groups will you purchase? Why?
b) What is the expected genetic superiority of the selected cows above random selection in this situation?

2. Assume a trait has heritability of .30, repeatability of .50, and phenotypic standard deviation of 400. The following information is available on 3 animals:

Animal	Number of Records	Average of Records
A	2	+800
B	1	+1000
C	0	--

Predicted genetic value of C's sire is +1200 with r_{TI} = .95, and estimated transmitting ability of C's dam is -200 with r_{TI} = .60.

a) Suppose you want to select for additive genetic value. In what order would you choose the animals to furnish herd replacements?
b) Suppose instead that you want to select for the sum of four future records. (Assume all animals would survive for four more records.) In what order would you cull?

3. The trait for the following problem can be measured on both males and females, has heritability of .36, and the variance of single records is 1600. A breeder is considering two selection programs. All selection is for additive genetic value.

Program 1: Males will be chosen on the basis of their own single record. Selection will be for the top 1 of 15 born each year. The generation interval will be 2 years. Females will be chosen on the basis of a single record of their sire and a single record of their dam. The generation interval will be 4 years. Selection will be for the best 18 of 20.

Program 2: Males will be chosen on the basis of records of 10 paternal half-sib progeny. Selection will be for the top 1 of 15 (possible by using artificial insemination). The generation interval will be 3½ years. Females will be chosen on the basis of a single record of their sire and a single record of their dam. Selection will be for the best 18 of 20, and the generation interval will be 4 years.

a) Which selection program should the breeder use to maximize the additive genetic value of the herd? (Show all calculations.)
b) Suppose that you could buy random progeny of the first generation of selection to start your own herd. Which selection program would you hope the breeder was following? (Show all calculations.)

4. The trait measured in this problem has heritability = .50, repeatability of records on the same animal = .75, and phenotypic variance = $(2000)^2$. Two full sibs (A and B) have records X_A = 3000 and X_B = -1500.

a) Predict the additive genetic value of A and of B.
b) What is the r_{TI} for predicting the additive genetic value of A?
c) What is the probability that the additive genetic value of A is greater than 400? (If unable to work a and b, assume $\hat{G}_A$ = 1200 with r^2_{TI} = .50.)
d) A future record of A is predicted by the selection index to be $\hat{R}_A$ = (8/15)(3000) + (1/15)(-1500) = 1500, and a future record of B is predicted to be $\hat{R}_B$ = (1/15)(3000) + (8/15)(-1500) = -600. The selection index for difference in predicted future records of A and B is:

$$I = \widehat{R_A - R_B} = [(8/15) - (1/15)](3000) + [(1/15) - (8/15)](-1500) = 2100 .$$

What is the probability that a future record of A, R_A, will exceed a future record of B, R_B, by 1000 or more? The r^2_{TI} is .311.

Final Exam, December 1975

1. For a trait, the following is known:

heritability = .40 ;
repeatability = .60 ;
phenotypic variance = $(90)^2$.

Records and inbreeding coefficients are available on the following unrelated animals.

Animal	Average Record	Number of Records	Inbreeding
A	+100	1	0
B	+100	1	.25
C	+50	1	0
D	-100	2	0
E	-100	1	0

a) Predict the additive genetic value for the five animals and rank them.
b) Predict a future record for each of the five animals and rank them.
c) What is the probability a future record of animal E will exceed +50?

2. You are selecting for traits 1 and 2 with an index:

$$I = 6X_1 + 5X_2$$

where X_1 and X_2 are records for traits 1 and 2 measured on the animal being evaluated.

Assume $\sigma^2_{X_1}$ = 100, $\sigma^2_{X_2}$ = 400, $\sigma_{X_1X_2}$ = 40, h_1 = .25, h^2_2 = .49, and $r_{g_1g_2}$ = -.50.

Assume the selection intensity factor is D = 1.5.

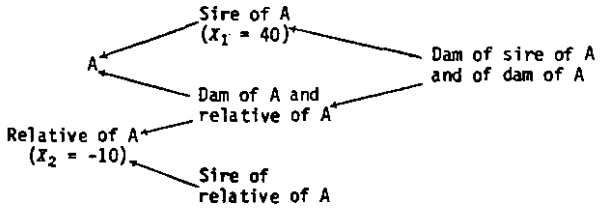
a) Calculate the expected genetic response per generation for trait 1.
b) Calculate the expected genetic response per generation for trait 2.
c) Suppose the economic values per unit are $6 for trait 1 and $5 for trait 2. What is the expected economic gain per generation?

3. Predict the value of animal A for:

$$T = 6G_{D_A} + G_{M_A}$$

where G_{D_A} and G_{M_A} are the additive genetic values of A for direct genetic effects and maternal genetic effects. Records are available on the sire ($X_1 = +40$) and another relative ($X_2 = -10$) as shown in the diagram.

$\sigma^2_{X_1} = 100$, $\sigma^2_{X_2} = 100$, $\sigma_{X_1X_2} = 0$, $\sigma^2_{G_D} = 40$, $\sigma^2_{G_M} = 20$, and $\sigma_{G_DG_M} = -20$.



4. You have decided since all your calves have the same breed and age of dam that the proper model to describe weaning weight of a population of beef cattle is:

Weaning Weight = Constant + Sex Effect + Herd Effect + Sire Effect + Residual Effect

which can be written in the usual notation as

$$y_{ijkl} = \mu + a_i + h_j + s_k + w_{ijkl} .$$

Assume sex and herd effects are fixed but unknown effects. The sire effects (sire transmitting abilities) are random effects with variance $\sigma^2_s = \frac{1}{4}\sigma^2_G = 50$. The variance of residual effects is $\sigma^2_w = 950$.

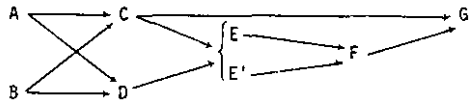
The following set of records is available:

(i) Sex	(j) Herd	(k) Sire	(l) Calf	(y_{ijkl}) Weaning Weight (kg)
1	1	1	1	200
1	1	1	2	240
1	1	1	3	220
1	1	1	4	180
2	1	1	5	200
1	1	2	1	260
2	1	2	2	240
2	1	2	3	220
1	1	3	1	200
2	1	3	2	220

Set up (numerically) the best linear unbiased prediction equations to predict the effects of sires 1, 2, and 3 and to estimate the fixed effects for which adjustments are needed.

First Exam, October, 1976

1. In the following pedigree, assume E and E' are identical twins (in a species where sex reversal is possible). Animals A and B are unrelated.



a) Show the additive relationships among all eight animals.

b) What is the *inbreeding coefficient* for C, for E, for E', and for G?

c) Calculate the dominance relationship between A and D and between C and D.

d) Suppose an unrelated animal is mated with E to produce another progeny, H. What will be the dominance relationship between C and H?

2. A dairyman believes that the first milk record of a cow is more important than her second record which is more important than her third record. Therefore, when he finds the average of three records, he enters the third record once, the second record twice, and the first record four times, and divides by seven, i.e., for cow i:

$$\bar{X}_i = \frac{X_{i1} + X_{i1} + X_{i1} + X_{i1} + X_{i2} + X_{i2} + X_{i3}}{7}$$

Assume the X_{ij} have been measured as deviations from the population mean so that $X_{ij} = G_i + PE_i + TE_{ij}$.

Assume the phenotypic variance is $(2000\ \text{lb})^2$, the genetic variance is $(1000\ \text{lb})^2$, and the variance of permanent environmental effects is $(900\ \text{lb})^2$. Note the phenotypic variance is $(2000\ \text{lb})^2 = 4{,}000{,}000\ \text{lb}^2$.

a) What is the variance of the "average" the dairyman is using?

b) What is the covariance of that "average" with the *genetic value* of the cow?

The usual average is $\bar{X}_i = (X_{i1} + X_{i2} + X_{i3})/3$.

c) What is the variance of the usual average?

d) What is the covariance of the usual average with the *producing ability* of the cow, $A_i = G_i + PE_i$?

3. Given two equivalent models:

(1) $P_i = \mu_P + G_i + E_i$

where $\mu_G = 0$, $\mu_E = 0$, $\mu_P \neq 0$, and with no covariance between G_i and E_i;

(2) $P_i = \mu + g_i + e_i$

where $\mu_g \neq 0$, $\mu_e \neq 0$, $\mu \neq 0$, and with no covariance between g_i and e_i.

Assume that $\text{Var}(G_i) = \text{Var}(g_i) = \sigma^2_G$ and $\text{Var}(E_i) = \text{Var}(e_i) = \sigma^2_E$.

a) Show that σ^2_P is the same for models (1) and (2).

b) If $\sigma^2_G = \sigma^2_{10} + \sigma^2_{01} + \sigma^2_{11} + \sigma^2_{20} = 30 + 20 + 10 + 40$ and $\sigma^2_E = \sigma^2_{PE} + \sigma^2_{TE} = 20 + 80 = 100$, what is the heritability for this trait; what is the repeatability for this trait?

4. A logical way to estimate the environmental covariance between relatives is to obtain two groups of relatives such that in one group an environmental covariance has a chance to be expressed (for example, full brothers raised in the same litter of pigs) and such that in the other group there is no chance for an environmental covariance to be expressed (for example, between full brothers separated at birth and raised in different litters). Suppose the following phenotypic covariances have been computed for weaning weight in pigs:

Covariance between full sibs in same litter	= 5.625
Covariance between full sibs raised in different litters	= 2.625
Covariance between paternal half-sibs in different litters	= 1.000
Covariance between identical twins in same litter	= 11.000
Covariance between identical twins raised in different litters	= 7.000
Variance of single records	= 25.000

a) What is the estimate of the environmental covariance between full sibs raised in the same litter?

b) What is the estimate of the environmental covariance between identical twins raised in the same litter?

Assume $\sigma^2_G = \sigma^2_{10} + \sigma^2_{01} + \sigma^2_{11}$.

c) Estimate σ^2_{01}. (Hint: $\hat{\sigma}^2_{11} = 1.$)

d) Estimate σ^2_{10}.

Second Exam, November, 1976

1. Do this question.

Given: $h^2 = .40$; $r = .55$; $\sigma^2_{10} = 400$; $\sigma^2_X = 1000$.

a) The following unrelated animals are to be ranked for additive genetic value. Their records and records of their relatives are given.

	Single Records		
Animal	Own	Sire	Dam
A	20	--	--
B	--	-15	10
C	5	-10	--
D	-10	--	20
E	--	--	30
F	5	15	--

b) What would you expect the superiority in additive genetic value to be if two of these six animals are selected?

2. Do this question.

a) The selection index procedure has been used to evaluate an animal.

$I = 18$; $r^2_{TI} = .84$; $r = .60$; $h^2 = .36$; $\sigma^2_T = (30)^2$;

$\sigma^2_X = (50)^2$.

What is the probability the true value of the animal is greater than 12?

b) Given: $\sigma^2_G = (30)^2$; $\sigma^2_X = (50)^2$; $r = .60$; $h^2 = .36$;

$E(X_1) = 0$.

The index for predicting a future record of an animal from a record of the sire, X_1, is $I = \frac{1}{2}h^2X_1$. Suppose for an animal that $I = 8$. What is the probability that a future record of the animal will exceed 48? What is the probability a future record will be less than zero?

3. Do either this question or question 4.

You are asked to compare two proposed boar testing procedures. Selection is for additive genetic value. Ten progeny are to be measured for gain from 21 days to 60 days. Assume 100 litters are available each year. Further assume that all litters have five pigs, that the gains are adjusted for sex and for age of the sow, and that all sows are unrelated to each other and to the boars.

Procedure A is to measure two pigs from each of five litters (20 boars can be tested). The other three pigs from each litter are not measured.

Procedure B is to measure five pigs from each of two litters (50 boars can be tested).

Assume the additive genetic variance is $.30\sigma^2_X$, the environmental covariance between full sibs is $.25\sigma^2_X$, and the environmental covariance between paternal half-sibs is zero.

Note that the weighting factors will be the same for all five litters in Plan A and similarly the weighting factors will be the same for both litters in Plan B. Note that selection intensity will not be the same for plans A and B.

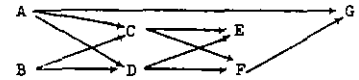
4. Do either this question or question 3.

Since the progeny of inbred parents resemble one another and their parents to a greater degree than those from noninbred parents, inbreeding is sometimes used for generating genetic uniformity.

A bull stud decides to progeny test 10 unrelated inbred bulls ($F = \frac{1}{4}$). They obtain 20 progeny per bull. Assume the dams are unrelated to each other and to the bulls. The bull stud plans to select the best two of these bulls based on progeny records (single record on each progeny).

Compare this procedure with testing 10 noninbred bulls. The assumptions are the same except that $F = 0$. Selection is for additive genetic value. Phenotypic variance = $(20)^2$, additive genetic variance = $(10)^2$, and variance of producing ability = $(15)^2$.

1. Animals A and B are unrelated. Both have inbreeding coefficients of .50. Mating is done as follows:



a) Show the additive relationships among the seven animals. Note especially the relationship between A and G and also that between F and G.

b) What are the inbreeding coefficients for the seven animals?

2. A researcher has calculated the covariance between maternal half-sibs to be 25. From the literature, he has estimates that the direct additive genetic variance for the trait is 40, the maternal additive genetic variance is 20, and the direct maternal genetic covariance is -10. The phenotypic variance is 100.

a) What is the heritability for the direct trait?

B) What is the heritability for the maternal trait?

c) What is the environmental covariance between maternal half-sibs?

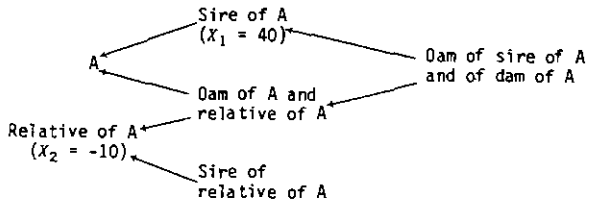
d) How much of the environmental covariance is not associated with genetic effects?

3. Predict the value of animal A for:

$$T = 6G_{D_A} + G_{M_A}$$

where G_{D_A} and G_{M_A} are the additive genetic values of A for direct genetic effects and maternal genetic effects. Records are available on the sire ($X_1 = +40$) and another relative ($X_2 = -10$) as shown in the diagram.

$\sigma^2_{X_1} = 100$, $\sigma^2_{X_2} = 100$, $\sigma_{X_1X_2} = 0$, $\sigma^2_{G_D} = 40$, $\sigma^2_{G_M} = 20$, and $\sigma_{G_DG_M} = -20$.



4. You have decided since all your calves have the same breed and age of dam that the proper model to describe weaning weight of a population of beef cattle is: ,

Weaning Weight = Constant + Sex Effect + Herd Effect + Sire Effect + Residual Effect

which can be written in the usual notation as

$$y_{ijkl} = \mu + a_i + h_j + s_k + w_{ijkl} .$$

Assume sex and herd effects are fixed but unknown effects. The sire effects (sire transmitting abilities) are random effects with variance $\sigma_s^2 = \frac{1}{4}\sigma_G^2 = 50$. The variance of residual effects is $\sigma_w^2 = 950$.

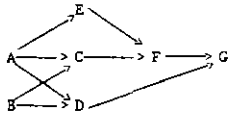
The following set of records is available:

(i) Sex	(j) Herd	(k) Sire	(l) Calf	(y_{ijkl}) Weaning Weight (kg)
1	1	1	1	200
1	1	1	2	240
1	1	1	3	220
1	1	1	4	180
2	1	1	5	200
1	1	2	1	260
2	1	2	2	240
2	1	2	3	220
1	1	3	1	200
2	1	3	2	220

Set up (numerically) the best linear unbiased prediction equations to predict the effects of sires 1, 2, and 3 and to estimate the fixed effects for which adjustments are needed.

First Exam, October, 1977

1. A breeder has been told that inbreeding will result in more uniform progeny. He decides to mate sire A with inbreeding coefficient of 1/2 to dam B which has inbreeding coefficient of 1/4. The parents of A are unrelated to the parents of B. A breeding plan as shown in the diagram is proposed. What will be the inbreeding coefficients of the 7 animals? What are the additive relationships of A with C, E, F, and G?



2. Environmental causes of likeness between relatives generally exist only if the relatives are kept together. Thus, one way of estimating the environmental covariance is from the difference between the covariance of a pair of relatives kept together and the covariance of the same pair of relatives kept apart.

The following six covariances have been computed.

Covariance (full sibs kept together)	= 44
Covariance (full sibs kept apart)	= 29
Covariance (paternal half sibs kept together)	= 17
Covariance (paternal half sibs kept apart)	= 11
Covariance (parent and progeny kept together)	= 24
Covariance (parent and progeny kept apart)	= 24

Additive genetic variance, additive by additive genetic variance, and dominance genetic variance as well as possible environmental covariances are to be estimated by you.

3. Known:

phenotypic variance	= 80
repeatability	= .60
heritability (narrow sense)	= .40
heritability (broad sense)	= .40

X_1 is a record on an animal.
X_2 is another record on the animal.
X_3 is a record on a full sib of the animal.
X_4 is a record on another full sib of the animal.

One breeder (A) uses as an index:

$$I_A = X_1 + X_2 + (X_3 + X_4)/2$$

Another breeder (B) uses as an index:

$$I_B = (X_1 + X_2)/2 + (X_3 + X_4)/4$$

a) Whar is the covariance berween X_1 and X_2?
b) The covariance between X_1 and X_3?
c) The covariance between X_2 and X_4?
d) The covariance between X_3 and X_4?
e) Which index will spread out predictions the most?
f) Give a numerical answer to (e).
g) Will rhere be any difference in ranking of animals if I_A is used as compared to using I_B?

4. Although impossible, assume a molecular geneticist can measure precisely additive, dominance, and additive by additive genetic effects. Further assume a cytogeneticist can identify exactly the origin of each gene at each locus. For a pair of Cornellius wombatus, the obscure species with five loci which has an affinity for adverse weather and ordinarily resides only in lab exercise #3, the two geneticists have established the following:

Molecular geneticist:

Additive gene effects associated with:

$a_1 = 4$	$b_1 = 1$	$c_1 = 2$	$d_1 = -2$	$e_1 = 0$
$a_2 = 2$	$b_2 = 0$	$c_2 = -1$	$d_2 = -1$	$e_2 = 0$
$a_3 = 0$	$b_3 = -1$	$c_3 = -1$	$d_3 = 3$	$e_3 = 0$

Dominance genetic effects associated with:

$a_1a_2 = -2$	$e_1e_3 = 4$	
$a_1a_1 = 1$	$e_1e_1 = 0$	all others = 0
$a_2a_2 = 1$	$e_2e_2 = 0$	

Additive by additive genetic effects associated with:

$b_1c_1 = -3$	$c_3e_1 = 4$	all others = 0
$b_1c_2 = 3$	$c_3e_3 = -4$	

Cytogeneticist:

Genotype of animal A: a_1a_3 b_1b_3 c_1c_1 d_1d_3 e_1e_3

Genotype of animal B: a_1a_2 b_1b_2 c_2c_3 d_2d_3 e_1e_3

Your task: Complete the following table

	Genetic Value			
Animal	Additive	Dominance	Additive by Additive	Total
A:				
B:				

Second Exam, November, 1977

1. An animal breeder (A) has devised a breeding plan which he feels is optimum for a trait. He can select the best 10% of males and best 80% of females with correlations between rrue and predicted genetic values of .90 and .45, respectively. The generation interval for males is 6 years and for females is 5 years. The heritability of the trait is .30 and repeatability is .45.

Another breeder (B) thinks her plan is optimum for another trait. She can select the best 20% of males and best 70% of females. The correlations with true genetic value are .85 and .50, and the generation intervals are 4 years and 6 years. The trait has heritability of .25 and repeatability of .50.

a) How much gain per year can breeder A expect?
b) How much gain per year can breeder B expect?
c) Which breeder (A or B) can expect to make the most progress? Why?

2. The transmitting abilities (one-half additive genetic value) of the following 3 sires are to be predicted. The 3 sires are unrelated to each other. All of their mates are unrelated to each other and unrelated to the sires. Heritability = 1/2; repeatability = 3/4; phenotypic variance = $(2{,}000)^2$.

Sire	Information available
A	10 progeny with one record each average +500.
B	5 progeny: first records average +400, second records average +600.
C	1 progeny with 1 record of +1000. 2 progeny, each with 4 records which average +500.

3. In herd 1, five daughters of sire A average 500 lb more than five daughters of sire B; i.e., $\bar{X}_{A_1} - \bar{X}_{B_1} = 500$ lb.

In herd 2, 10 other daughters of sire A average 200 lb less than 10 daughters of sire C; i.e., $\bar{X}_{A_2} - \bar{X}_{C_2} = -200$ lb.

Assume all daughters have unrelated dams. Heritability = 1/4; repeatability = 1/2; phenotypic variance = $(2{,}000\text{ lb})^2$.

a) Predict the difference in additive genetic value of sires B and C; i.e., $G_B - G_C$.

b) What is the correlation between $G_B - G_C$ and the prediction of $G_B - G_C$ from part a).

c) If the difference between G_A and G_B is to be predicted, should $\bar{X}_{A_2} - \bar{X}_{C_2}$ be used. If yes, indicate why and how without solving any equations.

4. A cow has two records that averaged +.4% in fat percentage. What is the probability that

a) the next record will be greater than .4%?

b) greater than 0.0%?

What is the probability rhat the average of five future records will be greater than

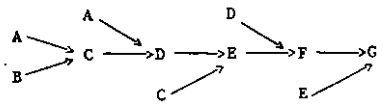
c) +.4%?

d) 0.0%?

Heritability = 1/2; repeatability = 6/10; phenotypic variance = $(.5\%)^2$.

Final Exam, December, 1977

1. The matings, indicated by the arrow diagram, are proposed as a way of building up inbreeding.



16 points a) Compute all additive relationships.
b) Which animals will be inbred, and whar are the inbreeding coefficients?

2 points c) Suppose that A and B are breeds rather than individual animals. The diagram then describes a crossbreeding pattern. Can the usual relationship table be used to indicate the fraction of genes each parent breed contributes to each generation? For example, what fraction of genes does breed A contribute to G?

2 points d) Derive a general formula to describe the contribution of breed A to progeny of the nth generation if the system of mating described in the diagram is continued.

NOTE THAT THIS IS A DOUBLE CREDIT PROBLEM

2. The following records are available:

Milk (pounds)	Fat (pounds)
Average of 20 half-sib daughters of sire A = +1000	One record of a full sister of sire A = +30
Average of 5 full-sib daughters of sire B = +1200	One record of dam of sire B = +20
	Average of 5 paternal half sisters of sire B = +40
One record of daughter D of sire C = +800	Average of 10 paternal half-sisters of sire C = +10
One record of dam of animal D = +1600	

The following information is also known:

	Milk	Fat
Phenotypic variance	$(2500)^2$	$(100)^2$
Heritability	.20	.25
Repeatability	.45	.50
Environmental covariances		
Between full sisters	$.16(2500)^2 = (1000)^2$	$.16(100)^2 = (40)^2$
Between half sisters	$.00(2500)^2 = 0$	$.00(100)^2 = 0$
Between daughter and dam	$.04(2500)^2 = (500)^2$	$.04(100)^2 = (20)^2$

32 points a) Rank sires A, B, and C for additive genetic value for milk yield.
b) Rank sires A, B, and C for additive genetic value for fat yield.

6 points c) Suppose fat has 10 times as much value per standard deviation as milk. The value of milk is $.05 per pound. Rank sires A, B, and C for their additive genetic economic value for milk and fat.

2 points d) What information do you need to improve the accuracy of the economic ranking if the same records are used?

3. A specialist in animal breeding has provided the following information to a bull stud:

	Prediction of Genetic Value		Prediction of record of a furure daughter	
Bull	Index value	Correlation with Genetic Value	Index value	Correlation with Record
A	2000	.50	1000	.125
B	800	1.00	400	.250
C	600	1.00	300	.250
D	-1000	.80	-500	.200

Phenotypic variance = $(2000)^2$ Heritability = .25

The bull stud has employed you to interpret such results. You decide to illustrate your knowledge (hopefully) by showing the manager that you can tell the bull stud the probability that a bull has genetic value greater than +1000 and thar you can tell a herd owner the probability that a record of a daughter will be +1500 or greater.

12 points
a) What is the probability that the genetic value of A is greater than +1000?
b) What is the probability that the genetic value of B is greater than +1000?

6 points c) What is the probability that a record of a daughter of B will be greater than +1500?

After you have done this, the manager also asks:

2 points d) What is the probability that a record of a daughter of B will exceed a record of a daughter of C by 500 or more?

4. A breeder asks you which of two selection methods you would recommend.

Merhod A: Index = $.9X_{1,\ own} - .4X_{2,\ own}$

Method B: Index = $.9X_{1,\ own} - .2X_{2,\ dam}$

$X_{1,own}$ and $X_{2,own}$ are traits 1 and 2 measured on the animal. $X_{2,dam}$ is trait 2 measured on the dam. Trait 2 is measured at a mature age while trait 1 is measured early in life. Thus, the generation intervals are 5 years for Method A and 3 years for Method B for both males and females. Assume selection intensities are equal for males and females. Males and females are selected by the same index. The situation in beef cattle where trait 1 would be yearling weight and trait 2 would be mature weight might be similar to this.

The phenotypic and genetic variances and covariances are shown below (variances are on diagonals and covariances on the off-diagonals).

Phenotypic	X_1	X_2
X_1	400	120
X_2	120	100

Genetic	G_1	G_2
G_1	121	55
G_2	55	36

The economic values are: for trait 1, \$3/unit, and
for trait 2, -\$1/unit

so that $T = \$3(G_{1,animal}) - \$1(G_{2,animal})$

3 points a) What is σ_T^2 ?

14 points
b) What are ΔG_1 , ΔG_2 , and ΔT for Method A?
c) What are ΔG_1 , ΔG_2 , and ΔT for Method B?

3 points d) Which method would you recommend to the breeder and why?

5. This is a bonus problem

(5 points) Suppose the record on an animal (x) is influenced by the maternal genetic ability of its dam (x') and by the value of the genotype of the fetus (w_x) born at the start of the record. The model for the record can be written as:

$$P_x = g_x + m_{x'} + f_{w_x} + e_x$$

where: g_x is the direct additive genetic value of x for the trait;

$m_{x'}$ is the maternal additive genetic value of x';

f_{w_x} is the fetal additive genetic value of w_x; and

e_x is the total of all other environmental effects.

The variances and covariances of the direct, maternal, and fetal additive genetic and of e effects of the same animal are:

	g_x	m_x	f_x	e_x
g_x	40	-5	-10	0
m_x		20	5	0
f_x			30	0
e_x				120

The diagonals are the variances and the off-diagonals are the covariances.

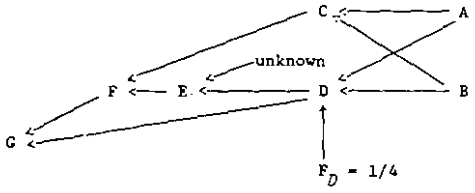
a) Develop a general symbolic expression in terms of relationships and genetic variances and covariances for the covariance between animals x and y with dams x' and y' and with records started by birth of fetuses w_x and w_y.

b) Calculate the covariance between records of two full sibs both of which were mated back to their fathers.

First Exam, October, 1978

(25 points)

1. The following arrow diagram describes a set of matings a breeder has made. Only E has a parent that was not related to the original parent pair, A and B. Note that $F_D = 1/4$ is also known. Compute the complete additive relationship table for the 7 animals.



(20 points)

2.

	A	B	C	D	E	F
					A - B	C - D
A	$1\frac{1}{4}$	0	$\frac{1}{2}$	0	$\frac{5}{8}$	$\frac{1}{4}$
B	0	$1\frac{1}{2}$	0	$\frac{3}{8}$	$\frac{3}{4}$	$\frac{3}{16}$
C	$\frac{1}{2}$	0	$1\frac{5}{8}$	0	$\frac{1}{4}$	$\frac{13}{16}$
D	0	$\frac{3}{8}$	0	$1\frac{3}{8}$	$\frac{3}{16}$	$\frac{11}{16}$
E	$\frac{5}{8}$	$\frac{3}{4}$	$\frac{1}{4}$	$\frac{3}{16}$		
F	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{13}{16}$	$\frac{11}{16}$		

Given the partially completed table of additive relationships:

a) What is the inbreeding coefficient of E?

b) What is the additive relationship between E and F?

c) What is the dominance relationship between E and F?

(25 points)

3. The following tables of additive and dominance relationships are known:

Additive

	A	B	C	D	E	F
A	1	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
B	0	1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
C	$\frac{1}{4}$	$\frac{1}{4}$	1	$\frac{1}{2}$	$\frac{3}{8}$	$\frac{3}{8}$
D	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	1	$\frac{3}{8}$	$\frac{3}{8}$
E	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$1\frac{1}{4}$	$\frac{5}{8}$
F	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{5}{8}$	$1\frac{1}{4}$

Dominance

	A	B	C	D	E	F
A	1	0	0	0	0	0
B	0	1	0	0	0	0
C	0	0	1	$\frac{1}{4}$	--	--
D	0	0	$\frac{1}{4}$	1	--	--
E	0	0	--	--	1	--
F	0	0	--	--	--	1

The following is also known:

Phenotypic variance = 100,
Heritability (narrow sense) = .32,
Repeatability = .60,
Dominance genetic variance = 16, and
Additive by dominance genetic variance = 8 .

a) What is the covariance between records of relatives C and D?

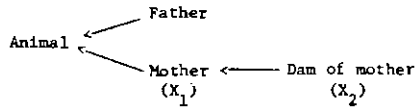
b) What is the covariance between records of relatives A and C?

c) What is the variance of single records of animals such as A?

d) What is the variance of single records of animals such as E?

(30 points)

4. The following index has been proposed to select for additive genetic value of an animal when a record on the animal's mother (X_1) and also a record on the dam of the mother (X_2) are available.



$I = 4X_1 + 2X_2$

Given $\sigma^2_X = \sigma^2_{X_1} = \sigma^2_{X_2} = 100$

$h^2 = .40$, $r = .60$

No environmental covariance between a mother and progeny.

a) What is the variance of true value?
b) What is the variance of the index?
c) What is the average squared prediction error of the index?
d) What is the r^2_{TI}?
e) How could you determine if this is the "best" index?

Second Exam, November, 1978

(50 points)

(40 points) 1. a) A breeder has asked you to rank the following six unrelated animals for additive genetic value.

Phenotypic variance = 100,000
Heritability = .20
Repeatability = .40
Environmental covariances:
Full sibs = 4,000
Paternal half sibs = 1,000
Parent and progeny = 2,000

Animal	Records available (difference from mean)
A	Record on A = 400
B	Record on B = 200
C	3 records on C ave. = 300
D	Records on 2 full sibs ave. = 600 (one record each)
E	Record on a paternal half sib = 1000
F	Sire of F has 5 records ave. = 300; Dam of F has 1 record = 400

(10 points) b) Indicate clearly without actually ranking the animals how you would predict a record for each of the six animals.

(25 points)

2. Selection indexes for two unrelated animals have been calculated. The animals were evaluated for genetic value.

Phenotypic variance = 1000
Genetic variances:
Additive genetic = 200
Dominance genetic = 100
Additive by additive genetic = 100
Permanent environmental variance = 150

Animal	I	r^2_{TI}
A	15	.36
B	10	.91

(10 points) a) What is the probability that the genetic value of A is greater than 10?

(5 points) b) What is the probability that the genetic value of B is greater than 10?

(5 points) c) What is the variance of the difference between the indexes, $I_A - I_B$?

(5 points) d) What is the variance of prediction error for predicting the difference in genetic values of animals A and B? The prediction is $I_A - I_B$.

(25 points)

3. Four sets of animals are available for sale. Each set includes two animals. Each animal has one record, but only the average for the set is known. Rank the groups according to their predicted additive genetic value.

The following information is also known:

Phenotypic variance = 100
Heritability (narrow) = .36
Repeatability = .60
Dominance genetic variance = 12
Environmental covariances:
Full sibs = 30
Paternal half sibs = 0
Unrelated = 0
Parent-progeny = 5

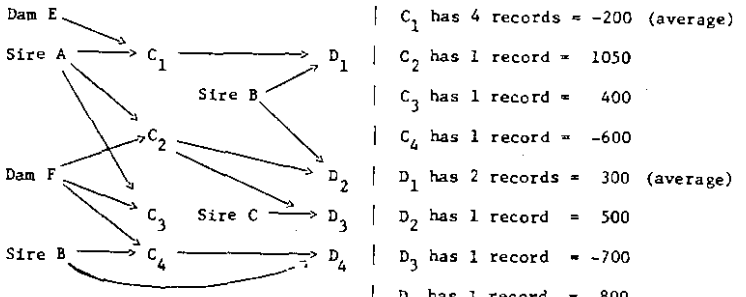
	Group	Average	Prediction
1)	2 full sibs	40	
2)	2 paternal half sibs	40	
3)	2 unrelated animals	40	
4)	parent and 1 progeny	40	

Remember to write down what information is available and what the true value is for each group.

Final Exam, December, 1978

(30 points)

1. The arrow diagram shows the relationships among all animals in a herd which have records.



C_1 has 4 records = -200 (average)
C_2 has 1 record = 1050
C_3 has 1 record = 400
C_4 has 1 record = -600
D_1 has 2 records = 300 (average)
D_2 has 1 record = 500
D_3 has 1 record = -700
D_4 has 1 record = 800

Phenotypic variance = $(400)^2$, heritability = .50, repeatability = .60. The environmental covariance between animals with the same sire is 200.

a) Show numerically how to set up the selection index equations to predict the additive genetic value for each of the animals which have records, i.e., C_1, C_2, C_3, C_4, D_1, D_2, D_3, and D_4 .

b) Sire C does not have a record. Can the additive genetic value of Sire C be predicted? If yes, what records would be used and how.

(30 points) 2. A breeder is convinced two traits (A and B) have economic importance and that he should select jointly for both of them. He is currently selecting on only one trait (A) which is measured earlier in life than the second trait (B). If both traits are used in the selection index, the generation interval will be increased by 2 years as shown in the table. The fractions selected for breeding for the two plans will be as shown:

Traits used in selection	Generation interval	Fraction selected out of survivors: Males	Females
A	3 yr	.10	.40
A and B	5 yr	.15	.60

The variances and covariances are:

	Phenotypic	
Trait	A	B
A	$(400)^2$	0
B	0	$(10)^2$

	Genetic	
Trait	A	B
A	$(200)^2$	100
B	100	$(4)^2$

Assume the correct economic values are \$3 per unit of trait A and \$20 per unit of trait B; i.e., $T = 3G_A + 20G_B$. Compare the two selection procedures which are the same for both males and females:

First procedure is $I = bX_A$ (Note that b can be any number $\neq 0$.)

Second procedure is $I = b_A X_A + b_B X_B$ (b_A and b_B are to be S.I. weights.)

(20 points)

3. Two unrelated sires are to be evaluated.

Sire A, by some miracle, is completely inbred, $F_A = 1$.

Sire B is not inbred, $F_B = 0$.

Each sire has 20 progeny. The 40 mothers are unrelated and unrelated to the sires. The progeny averages are:

Sire	Progeny average (single records of 20 progeny)
A	400
B	400

Phenotypic variance = $(200)^2$, additive genetic variance = $(100)^2$, repeatability = 1/2, heritability = 1/4.

a) Predict the additive genetic value for each sire.
b) Calculate the correlation between true and predicted additive genetic value (r_{TI}) for each of the two sires. (You can leave the answers as r^2_{TI} .)

Now assume for c) and d) that the environmental covariance among records of progeny of sire B equals one-fourth the phenotypic variance.

c) What is the prediction of the additive genetic value of sire B?
d) What is the r_{TI} for sire B? (Or r^2_{TI} ?)

(20 points + 3?)

4. Calving difficulty is measured as a categorical trait. If the only two categories are (1) difficult, and (2) not difficult, the model is the same as for traits measured on a continuous scale. In this case the measurements are 1, for a difficult birth, and 0, for a not difficult birth. The sire of the calf (mate of the cow) can influence the birth through genes transmitted to the calf. There may also be a need to evaluate the sire of the cow for the ease which his daughters give birth. Thus, two predictions are needed: effect of sire of calf and effect of sire of cow.

Separate predictions should be made for each sex of calf but for this problem assume sex of calf is in the model as a fixed effect.

The model for this problem is

$$y_{ijkl} = \mu + x_i + s_j + m_k + w_{ijkl} \quad ,$$

where μ is a constant,
x_1 and x_2 are effects of female and male births,
s_j (j=1,2,3) is the effect of the sire of calf ($\sigma^2_s = .005$)
m_k (k=1,2,3) is the effect of the sire of cow ($\sigma^2_m = .002$)
w_{ijkl} is the residual effect ($\sigma^2_w = .160$), and
y_{ijkl} is the observation on the lth cow (1 if a difficult birth and 0 if a not difficult birth), ($\sigma^2_y = .167$) .

The following table summarizes a set of data. (A real set would probably have quite a different pattern.)

Number of difficult births ($y_{ijk.}$)

Sex(i)	MGS (k)	Sire (j) A	B	C	Sum
1	D	10	5	0	
	E	-	4	10	
	F	15	1	5	
2	D	10	6	4	
	E	-	10	15	
	F	20	2	5	
Sum					

Number of calvings (n_{ijk})

Sex(i)	MGS (k)	Sire (j) A	B	C	Total
1	D	30	40	10	
	E	-	60	70	
	F	50	10	40	
2	D	20	30	20	
	E	-	70	60	
	F	60	20	50	

a) Set up the equations needed to predict the effects of sires A, B, and C as sires of calves and the effects of maternal grandsires D, E, and F as sires of cows giving birth. (Do not attempt to solve the equations.)

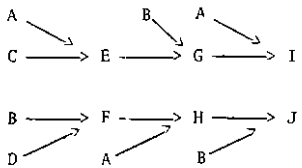
b) Bonus of 3 points (Not credited if you obtain 20 points on part a)

How many equations should be set up before constraint(s) are applied?
How many constraints are needed?
Write down a set of equations to solve the desired predictions.
(Do not attempt to solve the equations.)

A. S. 420
First Exam
October, 1979

30 points

1. A and B are full sibs whose parents are related by 50%. Animals A and B are the basis of a linebreeding program which has progressed as follows:



a) Which of the ten animals, A to J, are inbred?

b) What are the inbreeding coefficients?

c) Show the additive relationships among the ten animals.

d) Can you calculate the dominance relationship between E and F? If so, what is it?

40 points

2. You are to rank animals A, B, and C. Each has the same kind of available information, X_1 and X_2.

The selection index equations are:

$$400\ b_1 + 40\ b_2 = 155$$

$$40\ b_1 + 400\ b_2 = 65$$

The selection index is:

$$I = 3/8\ X_1 + 1/8\ X_2$$

Can you compute the following? If yes, what are the numerical values? If no, why not?

a) σ_I^2 : yes no σ_I^2 =

b) r_{TI}^2 : yes no r_{TI}^2 =

c) V(T-I): yes no V(T-I) =

d) Can you evaluate animal D which has only X_1 available? If so, how would you do it?

A friend thinks the index should be

$$I = 1/8\ X_1 + 3/8\ X_2$$

Can you compute the following? If yes, what are the numerical values? If no, why not?

e) σ_I^2 : yes no σ_I^2 =

f) r_{TI}^2 : yes no r_{TI}^2 =

g) Can you say which r_{TI}^2 (the first index or the second) will be larger? Why or why not?

h) If yes, how much larger?

30 points

3. The following were part of larger a_{ij}, d_{ij}, and c_{ij} tables.

a_{ij}

	A	B	C	D
A	1	0	1/2	1/2
B	0	1¼	0	5/8
C	1/2	0	1	1/4
D	1/2	5/8	1/4	0

d_{ij}

	A	B	C	D
A	1	0	1/4	1/4
B	0	1	0	0
C	1/4	0	1	0
D	1/4	0	0	1

c_{ij}

	A	B	C	D
A	.15	0	1/5	0
B	0	.15	0	0
C	1/5	0	.15	1/8
D	0	0	1/8	.15

The following variances are also known:

additive genetic variance	=	40
dominance genetic variance	=	20
additive by additive genetic variance	=	24
additive by dominance genetic variance	=	16
environmental variance	=	100

a) What is heritability of this trait?

b) What is repeatability of this trait?

c) What is the covariance between a record of A and a record of C?

d) What is the covariance between a record of A and a record of D?

e) What is the covariance between a record of C and a record of D?

f) What is the covariance between a record of A and another record of A?

g) What is the covariance between a record of B and another record of B?

A. S. 420
Second Exam
November, 1979

30 points

1. The variances and covariances needed to derive selection indexes using repeated records on an animal are given in the following table.

	Record				Covariances with	
	1st	2nd	3rd	4th	Future Record of Animal	Additive Genetic Value of Animal
1st	50	30	30	30	30	20
2nd	30	50	30	30	30	20
3rd	30	30	50	30	30	20
4th	30	30	30	50	30	20

a) What is heritability of the trait?

b) What is repeatability of the trait?

c) Predict a future record for the following four animals.

Animal A which has a first record of +15

Animal B which has two records, +10, +8

Animal C which has four records, the average is +5

Animal D which is the progeny of animal A.

2. 30 points

a) Your employer asks you as an expert in animal breeding to buy 3 animals for his herd. The only animals available are owned by a progressive farmer who has indexed his animals for additive genetic value (which is what you know you must select for).

However!! He will only sell one pen of animals and only the whole pen. You cannot select animals from more than one pen. What is your decision? Please state your reasons.

$h^2 = .25$ $r = .40$ $\sigma_x^2 = (2000)^2$

Pen 1		Pen 2		Pen 3		Pen 4	
I	r_{TI}	I	r_{TI}	I	r_{TI}	I	r_{TI}
100	.50	300	.80	0	.20	600	.40
0	.10	300	.10	0	.20	-300	.20
-100	.40	-600	.10	0	.60	-300	.40

b) If you were allowed to buy 4 animals but only the best one per pen what would be the expected additive genetic value of the selected animals?

10 points

3. Records on relatives of an animal have been used to predict the additive genetic value of an animal, the selection index is:

$$I = 27.2 \quad \text{with } r^2_{TI} = .69 \quad .$$

The same records have been used to predict a future record of the same animal; again the selection index is:

$$I = 27.2 \quad .$$

The computer, however, was not programmed to compute the r^2_{TI} for a future record.

The trait has heritability = .40 and repeatability = .60 .

a) Can the r^2_{TI} for predicting a future record be computed from the information given?
If no, why not?
If yes, do it and show the steps used in arriving at your answer!

b) Can the variance of prediction error for predicting a future record be computed from the information given?
If no, why not?
If yes, do it and show the steps used in arriving at your answer!

30 points

4. Selection indexes to predict future records for two unrelated animals, A and B, have been computed from records of relatives. (There is no covariance between the index for A and the index for B.)

heritability = .30

repeatability = .40

phenotypic variance = 100

$I_A = 25 \quad r^2_{TI_A} = .19$

$I_B = 15 \quad r^2_{TI_B} = .19$

a) What is the probability that the next record of animal A will exceed 25?

b) What is the probability the next record of animal A will exceed 15?

c) What is the probability the next record of animal B will exceed 25?

d) What is the probability the next record of animal B will exceed the next record of animal A?

Final Exam, December, 1979

25 points

1. X_A, X_B, X_C, X_D are measurements on an animal for traits A, B, C, and D. The phenotypic and genetic variances and covariances are:

Phenotypic variances and covariances

	X_A	X_B	X_C	X_D
X_A	200	-50	-20	-40
X_B	-50	100	10	20
X_C	-20	10	300	80
X_D	-40	20	80	400

Genetic variances and covariances

	G_A	G_B	G_C	G_D
G_A	20	10	-10	-20
G_B	10	30	10	20
G_C	-10	10	50	30
G_D	-20	20	30	100

A breeder has been using the index

$$I = 3X_A + 2X_B \quad .$$

You have discovered the correct economic value is

$$T = G_B - G_C + 2G_D \quad .$$

a) Calculate the expected superiority in T if the best 20 out of 50 animals are selected based on $I = 3X_A + 2X_B$ where traits X_A and X_B are measured on each animal.

b) Calculate the expected superiority in G_D if selection is based on $I = 3X_A + 2X_B$.

25 points

2. Within the foreseeable future it may be possible to obtain female progeny of a bull for which, in a genetic sense, the sire is both the mother and father. The process would be equivalent to joining a pair of random gametes (X-carrying only) of the sire and then incubating the zygote in a recipient female. This introduction applies to part c.
What would be the accuracy (r_{TI}) of predicting the additive genetic value of a sire from:

a) 20 progeny which result from mating the sire to 20 unrelated females.

b) 20 progeny which result from mating the sire to only one super-ovulated female with the fertilized ova transferred to recipient females.

c) 20 progeny for which the sire is genetically both the mother and father.

$h^2 = .25$

$r = .50$

$\sigma^2_p = (2000 \text{ lb})^2$

25 points

3. A breeder of Charolais cattle thinks age of the dam, sex of the calf and sire of the calf affect the weaning weight of the calf.

The model would be

$$y_{ijkl} = \mu + a_i + x_j + s_k + w_{ijkl}$$

where a_i is the fixed effect of the ith age of dam,

x_j is the fixed effect of the jth sex of calf,

s_k is the random effect common to calves of sire k with mean zero, and variance $\sigma_s^2 = 600$, and

w_{ijkl} is a random effect which includes the other genetic effects on the calf and environmental factors with mean zero and variance $\sigma_w^2 = 6000$.

The following data adjusted for days at weaning to a 205-day basis have been collected.

Calf	Age of dam (yr)	Sex of calf	Sire of calf	Weaning wt. (lb)
A	3	M	Grandiose	400
B	3	F	Grandiose	450
C	3	M	Grandiose	350
D	3	M	Bellicose	400
E	3	M	Bellicose	500
F	3	F	Comatose	300
G	3	F	Comatose	400
H	3	M	Comatose	500
I	3	F	Comatose	500
J	3	F	Comatose	400

a) Show numerically how you would jointly estimate the effects of age of dam and sex of calf and predict one-half the genetic values of the three sires.

b) How many constraints will be needed to obtain solutions?

25 points

4. Many animal breeders have difficulty with interpreting what has been estimated by solutions to least squares equations. As you know, the constraint(s) imposed on the equations in order to obtain solutions determine what has been estimated. A popular constraint is the sum to zero constraint; that is, the sum of the solutions for a particular kind of effect are forced to sum to zero.

Consider the model

$$y_{ij} = \mu + a_i + w_{ij}$$

where i = 1, 2, or 3, and a_i is the fixed but unknown effect of the ith age. y_{ij} is the record of animal j of the ith age. The w_{ij} are uncorrelated with variance σ_w^2 and mean zero. Let n_i be the number of records of animals of age i. Then the least squares equations with the sum to zero constraint are

$$\begin{aligned}
n_. \hat{\mu} + n_1\hat{a}_1 + n_2\hat{a}_2 + n_3\hat{a}_3 &= y_{..} \\
n_1 \hat{\mu} + n_1\hat{a}_1 + \lambda &= y_{1.} \\
n_2 \hat{\mu} + n_2\hat{a}_2 + \lambda &= y_{2.} \\
n_3 \hat{\mu} + n_3\hat{a}_3 + \lambda &= y_{3.} \\
\hat{a}_1 + \hat{a}_2 + \hat{a}_3 &= 0
\end{aligned}$$

Notice the new expression, λ, which has the official title of a LaGrange multiplier, is added to maintain the symmetrical form of the equations and to allow a solution to be obtained.

With this constraint when $n_1 = n_2 = n_3 = n$, $\hat{\mu} = \bar{y}_{..}$, $\lambda = 0$, $\hat{a}_1 = \bar{y}_{1.} - \bar{y}_{..}$, $\hat{a}_2 = \bar{y}_{2.} - \bar{y}_{..}$, and $\hat{a}_3 = \bar{y}_{3.} - \bar{y}_{..}$, where the dot notation denotes summation over that subscript and the bar denotes average; i.e., $\bar{y}_{1.} = y_{1.}/n_1$

Answer the following when $n_i = n$, and $\hat{a}_1 + \hat{a}_2 + \hat{a}_3 = 0$.

a) What function of the parameters (μ, a_1, a_2, and a_3) is estimated by $\hat{\mu}$?

b) By $\hat{a}_1$?

c) Can a_1 be estimated?

d) Can μ be estimated?

Hint: Try for $n_1 = n_2 = n_3 = 3$ if you have trouble with the idea of n observations.

ANIMAL SCIENCE 420

First Exam
October, 1980

1. You are to rank the following animals for additive genetic value for some trait and explain the reasons for your decisions. Different kinds of records are available for each animal, but a kind friend has worked out the selection index weights and the r_{TI}'s for each index.

	Available records			S. I. b's for			
Animal	X_1	X_2	X_3	X_1	X_2	X_3	r_{TI}
A	45	-30		7/15	2/15		.73
B			80			1/4	.35
C	-15		90	7/15		2/15	.73
D	40			1/2			.71
E		-20	20		1/4	1/4	.50
F		20			1/4		.35
G	30		-105	7/15		2/15	.73

2. The phenotypic variance for a trait is 100.
The covariance between a parent (X_1) and progeny (X_2) record is 20.
The covariance between true value (T) and X_1 is also 20.
The covariance between T and X_2 is 40.
The variance of T is also 40.

You decide to select on the basis of the index

$$I_{you} = X_1 + 2X_2 \quad .$$

Your friend prefers the index

$$I_{friend} = 5X_1 + 10X_2 \quad .$$

a) What is the variance of your index?

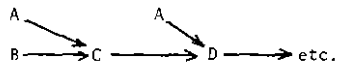
b) What is the variance of your friend's index?

c) What is the covariance between your index and your friend's index?

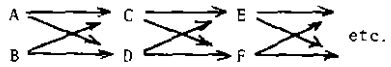
d) Which is the better index? Why?

3. Fathers are normally related to their daughters by 50%. Full brothers and sisters are also normally related by 50%. Matings of related animals result in inbreeding. Two systems of increasing inbreeding are:

1) to mate a sire back to his daughters, then his granddaughters, etc. The system starts:

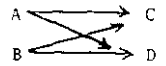


2) to make continuous full brother/sister matings. The system starts:



Which system will increase inbreeding the more rapidly? Explain!

4. Suppose the impossible happens and two completely inbred but unrelated animals are mated as shown.



The table of additive relationships is:

	A	B	C	D
A	2	0	1	1
B	0	2	1	1
C	1	1	1	1
D	1	1	1	1

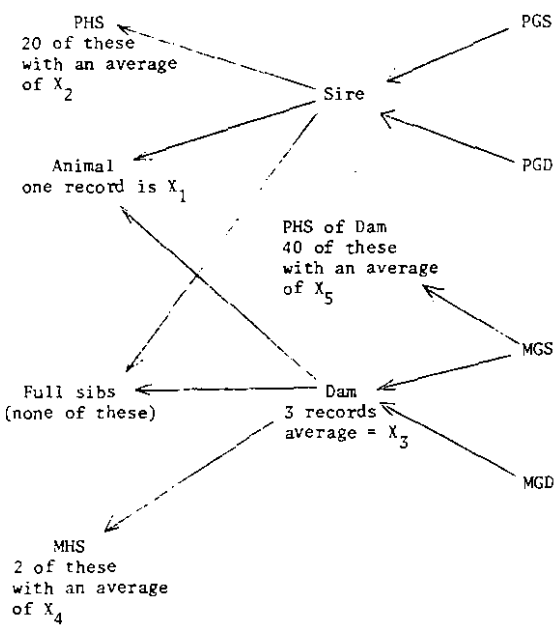
Suppose for some trait:

additive genetic variance = 50
dominance genetic variance = 20
additive by additive genetic variance = 30
permanent environmental variance = 20
temporary environmental variance = 80

a) What is heritability of this trait?
b) What is repeatability of this trait?
c) What is the covariance between a record of C and a record of D?
d) The covariance between a record of A and a record of D?
e) The covariance between a record of A and a record of B?
f) The covariance between two records of A?
g) The covariance between two records of C?

Second Exam, November, 1980

DIAGRAM PROBLEM 1



1. You are asked by a breeding organization to find the weights to be given to records of various relatives for predicting a future record of an animal. The organization has a computer to do the calculations. You only need to tell the computer programmer what must be computed. The diagram on the facing page shows the relationships among the animals and the records available. $h^2 = .50$, $r = .60$, $\sigma_x^2 = (2000)^2$, $\sigma_{01}^2 = .10(2000)^2$, the environmental covariance between full sibs is $.05(2000)^2$, the environmental covariance between maternal half-sibs is $.00(2000)^2$, the environmental covariance between paternal half-sibs is $.01(2000)^2$.

   A) You must show the programmer both symbolically and numerically what must be done. You do not have to do any arithmetic calculations.

   B) Show the programmer what must be changed if the animal has no record, but the average of single records of two full sibs is X_1 (i.e., replace the record of the animal with the average of 2 full sibs).

2. Selection decisions usually can be reduced to the comparison of pairs of animals--one to be culled and the other to be selected or to be compared with still another animal.

   For the following 4 situations, predict the difference in additive genetic value for the pair of animals involved.

   A) A has a record of +200.
      B has a record of -200.
      They are not related.
      Predict the difference between A and B.

   B) A has a record of +100.
      B has a record of -100.
      They are paternal half-sibs (same father, but different mothers which are unrelated).
      Predict the difference between A and B.

   C) Two full-sibs of A and B average +150.
      Predict the difference between A and B.

   D) The father of A and B has a record of +50.
      A has a record of -100.
      B does not have a record. The mothers of A and B are unrelated.
      Predict the difference between A and B.

   Heritability = .4, repeatability = .6, phenotypic variance = $(100)^2$,
   the environmental covariance between full sibs is $.10(100)^2$,
   the environmental covariance between paternal half-sibs is $.02(100)^2$.

3. A cow has a record of +4000 lb of milk.

   A) What is the probability that her next record will be +4000 lb or greater?

   B) What is the probability that the average of her next five records (after the first) will be +4000 or greater?

   These problems may have several steps. If you are not sure of the answer for a particular step, indicate symbolically the correct answer, make a reasonable approximation, and proceed with the problem.

   Assume heritability = .36
   phenotypic variance = $4{,}000{,}000(1b^2)$
   the covariance between records on the same animal = $2{,}000{,}000(1b^2)$

4. Perfect sexing of semen (i.e. into a part always producing males and a part always producing females) could increase the selection intensity factor for two of the paths of selection but may reduce selection for another path.

   E.g.

a)	cows to produce heifers	top 45% rather than top 90%
b)	cows to produce sons	top 1% rather than top 2%
c)	bulls to produce sons	top 5% for both cases
d)	bulls to produce heifers	top 20% when not separating semen top 25% for sexed semen because of semen loss in processing

   We will assume for purposes of this problem that sexing of semen is possible although currently that is not true.

   Assume the accuracies of evaluation for the four paths are for

   a) .58 (based on 2 records/cow)
   b) .61 (based on 3 records/cow)
   c) .90 (based on 50+ daughters per bull)
   d) .85 (based on 50 daughters per bull)

   Also assume the generation intervals are 5, 6, 8, and 6 years.

   Heritability is 25%

   A) Compare the expected genetic gain for perfect sexing of semen with use of regular semen.

   B) What two important factors must be considered before deciding whether to recommend sexing of semen?

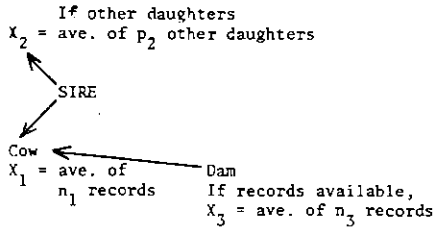
Final Exam, December, 1980

Maximum loss of 16 points on this problem.

This is a selection index problem.

1. A professor from a university in Buenos Aires wrote on November 4, 1980, that they intend to begin calculating estimated transmitting abilities (one-half of additive genetic value) of dairy cows. He wants a complete set of weighting factors for calculating the herdmate ETA of a cow. The word herdmate indicates that all records are expressed as differences from herdmate averages. In that way the records are corrected for management levels. Thus, the selection index procedure can be used.

   Since a complete table of weighting factors is unavailable, you must show the professor how the weighting factors can be determined for up to three kinds of records. The cow, herself, has at least one record. Her mother may have none up to several records. Her sire may have no other daughters up to several with one record each.



Assume heritability is .20
repeatability is .40
phenotypic standard deviation is 2000 lb

A) Indicate numerically how the weighting factors would be determined. (Just set up the equations numerically, the computer programmer can then calculate the weights for different n_1, p_2, and n_3.

B) Suppose the professor then writes back and asks how to modify the weights if there is dominance genetic variance and environmental covariances.

Dominance variance = $(.1)(2000)^2$
Environmental covariances
full sibs $(.06)(2000)^2$
paternal half sibs $(.04)(2000)^2$
parent and progeny $(.01)(2000)^2$

Indicate numerically what must be changed to determine the weighting factors.

16 points can be lost on this problem.

This is a selection index problem and not a mixed model problem.

2. A bull has 40 daughters (all with unrelated mothers) which average +2000.

The bull has an inbreeding coefficient of	.25
Heritability is	.40
Repeatability is	.60
Phenotypic variance is	1,000,000
Dominance variance is	100,000
Environmental covariances are	
Parent-progeny	50,000
Paternal half-sibs	10,000
Full sibs	200,000

Maximum loss out of 25 points

(9) A) What is the probability (chance) the transmitting ability of the bull (one-half of his additive genetic value) equals or exceeds the average of his first 40 daughters? (Complete all numerical computations.)

(3) B) What is the probability (chance) that the record of the 41st daughter will exceed the average of the first 40. (Set up numerically so that anyone could do the calculations, but you do not need to complete the calculations if the instructions are clear.)

(2) C) What is the probability (chance) that the next 40 daughters will average as much as the first 40? (The next 40 have unrelated mothers which are unrelated to the mothers of the first 40.) (Set up numerically.)

(2) D) Suppose the bull was mated back to the same cows and with much luck 40 more daughters were obtained. What is the probability (chance) that they would average as much or more than the average of the first 40? This problem violates some of the assumptions of the selection index procedure, but the S.I. procedure can be used. (Set up numerically.)

16 points can be lost on this problem.

3. Weights at various ages are the economically most important traits of beef cattle.

   A heavy birth weight (BW) may cause calving difficulty. If a calf is too small, it may not be strong enough to live.

   Weaning weight (WW) is the saleable product of the rancher.

   Yearling weight (YW) is an indication of weight at slaughter--the final marketable product.

   Mature weight (MW) must be maintained and thus heavy animals are more costly than smaller animals.

   Assume $T = -.03G_{BW} + .60G_{WW} + .40(G_{YW} - G_{WW}) - .05G_{MW}$

   The variances (diagonals) and covariances (off-diagonals) are:

	Phenotypic				Genetic			
	BW	WW	YW	MW	BW	WW	YW	MW
BW:	$(10)^2$	80	100	100	$(5)^2$	30	40	40
WW:		$(40)^2$	2000	2100		$(25)^2$	1000	900
YW:			$(100)^2$	5000			$(70)^2$	2500
MW:				$(120)^2$				$(65)^2$

   Compare the economic gain expected from the use of three indexes. (Assume an equal fraction selected and equal generation intervals.)

A) An approximate index (essentially heritability times economic value) based only on BW and WW (when choice of replacement heifers would be made).

$$I_A = -.0075BW + .24WW$$

B) An approximate index (if YW were available).

$$I_B = -.0075BW + .24WW + .20(YW - WW)$$

C) An approximate index at weaning time using the mother's mature weight (divided by two).

$$I_C = -.0075BW + .24WW - .015(\text{mother's MW}/2)$$

16 points can be lost on this problem.

This is a mixed model problem not a selection index problem.

4. The following set of 5 records is representative of a much larger set. To avoid a 16-point loss from your grade in this exam you must:

   A) write the model and equations to be solved to predict real producing ability (G + PE) for the three animals <u>which</u> <u>are unrelated</u>.

   Heritability = .40 Repeatability = .60 Phenotypic variance = $(40)^2$

   Animal 104, first record at 24 months of age = 360

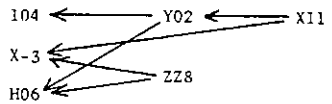
   Animal 104, second record at 36 months of age = 440

   Animal X-3, first record at 24 months of age = 420

   Animal H06, first record at 24 months of age = 380

   Animal H06, second record at 36 months of age = 400

   B) <u>If</u> the animals <u>are</u> <u>related</u> as shown in the diagram



   now indicate how the mixed model equations should be changed (numerically as far as you can).

Second Exam, November, 1982

1. The following records are available (all are adjusted for μ):

Animal 1; single record = +10
Animal 2; single record = +12 } average = +14
Animal 3; single record = +16

The three animals are full sibs (same sire and dam).

There is a high environmental correlation among full sibs for this trait, $c_{FS} = .40$. Heritahility for the trait is $h^2 = .20$; the phenotypic standard deviation is $\sigma_X = 5$.

a) Predict the additive genetic value of animal 1.

b) What is the accuracy of the prediction, r_{TI}?

The breeder discovers that there were really four full sibs in the group but animal 4 was not recorded for that trait.

c) Predict the record for animal 4.

2. Two bulls have been evaluated for additive genetic value:

$I_A = 1200$ with $r^2_{TI} = .36$

$I_B = 800$ with $r^2_{TI} = .84$

For this trait:

$h^2 = .25$

$r = .40$

$\sigma^2_X = (2000)^2$

a) What is the probability the additive genetic value of A is greater than 0?

b) What is the probability the additive genetic value of B is greater than 0?

3. Recent reports indicate that clones of animals can be obtained by splitting off cells of a developing blastocyst. These cells are genetically identical. Some of the cells can result in offspring which have a record. Others of the cells can be frozen, and after their clone mates have provided a "clone" test, can be thawed, grown in culture, and resplit to obtain more clones. At least, that is what is supposed to happen.

For the trait of interest:

$\sigma^2_{IO} = 40$

$\sigma^2_{PE} = 60$

$\sigma^2_{TE} = 100$

Assume that single records of 5 members of the clone are used to "prove" the clone.

a) What is the index for predicting a future record of one of the frozen cells?

b) What is the accuracy of predicting additive genetic value of one of the frozen cells?

c) What is the accuracy of predicting additive genetic value of one of the five of the clone that has a record?

4. A beef cattle ranch wants to select for increased weaning weight but also knows that increasing weaning weight may increase the size of its cows and thus increase maintenance costs.

Assume

$$T = (.30)G_{ww} - (.02)G_{mw}$$

There is a problem in measuring mature weight but assume it can be obtained reasonably well at 2 years of age.

The parameters are:

	Heritability	Variance: Genetic	Variance: Phenotypic
Weaning weight (ww)	.25	$(15)^2$	$(30)^2$
Mature weight (mw)	.49	$(63)^2$	$(90)^2$

The correlations are: genetic, .60; phenotypic, .40.

The index to predict T from an animal's own weaning weight and mature weight is: $I = .058286X_{ww} + .003429X_{mw}$.

The r_{TI} for this index is .4883.

Waiting to use the animal's own mature weight requires selection be delayed for over a year adding to the cost of maintenance and an increase in generation interval (assume 6 years).

One way to make the selection at weaning time would be to predict T from the animal's own weaning weight and the dam's mature weight.

a) Find the index to predict T from an animal's weaning weight and its dam's mature weight. (Assume generation interval is reduced to 5 years.)

b) What is the r_{TI} for this index?

c) If the intensity of selection is the top 30% for both indexes, and if the generation intervals are 5 years for a) and 6 years using the animal's own mature weight, what is the expected economic progress by using the index using the dam's mature weight (a) as compared to the index using the animal's own mature weight?

Final Exam, December, 1982

1. (30 points)

Given: $\sigma^2_{10} = 40$ $\quad C_{\text{full sibs}} = .125$

$\sigma^2_{01} = 20$ $\quad C_{\text{half sibs}} = .05$

$\sigma^2_{20} = 32$

$\sigma^2_{PE} = 8$

$\sigma^2_{TE} = 100$

a) Predict the genetic value for each of the following animals:

i) Animal A: Sire's record = 8, dam's record = 14

ii) Animal B: Average of single records of 5 otherwise unrelated progeny = 15

iii) Animal C: own record = 10, a full sib record = 5

iv) Animal D: own record = 12, $F_D = .25$

b) i) What is the r_{TI} for animal A?

ii) For animal B?

iii) For animal C?

iv) For animal D?

2. (30 points)

Trait A cannot be measured until 2 years of age.

Trait B cannot be measured until 3 years of age.

If selection is based on the animal's records for trait A and trait B, the generation interval would be 6 years.

If selection were based on the mother's records for trait A and trait B,

$$I_1 = \beta_1 X_{dam,A} + \beta_2 X_{dam,B} \quad ,$$

the generation interval would be only 4 years.

If selection were based on the animal's record for trait A and the dam's record for trait B,

$$I_2 = \beta_1^* X_{own,A} + \beta_2^* X_{dam,B} \quad ,$$

the generation interval would be 5 years.

Assume the top 50% would be chosen whether I_1 or I_2 were used.

Assume the overall economic genetic value is:

$$T = .3(G_{own,A}) - .1(G_{own,B}) \quad .$$

Phenotypic variances and covariances

	X_A	X_B	X_C
X_A	100	0	50
X_B	0	400	100
X_C	50	100	900

Genetic variances and covariances

	G_A	G_B	G_C
G_A	40	-20	10
G_B	-20	80	30
G_C	10	30	300

Variances and covariances involving trait C are to be used in b) for calculating correlated response.

a) Compare the annual economic gain expected from use of I_1 with that from the use of I_2.

b) Compare the expected yearly correlated response in trait C from use of I_1 with that from the use of I_2. If you did not find the weights for I_1 and I_2 use:

$$I_1^* = .07X_{dam,A} - .02X_{dam,B}$$

and $$I_2^* = .14X_{own,A} - .01X_{dam,B}$$ for your calculations.

3. (25 points)

The model for a record is:

$$y_{ijk\ell} = \mu + H_i + A_j + g_{ik} + p_{ik} + w_{ijk\ell}$$

where μ is a constant,

H_i is the fixed effect of herd i (management level),

A_j is the fixed effect of age j,

g_{ik} is the additive genetic value of animal k in herd i,

$(g_{ik} + p_{ik})$ is the real producing ability of animal k in herd i (p_{ik} = producing ability - additive genetic value), and

$w_{ijk\ell}$ is the residual (temporary environmental effect on record ℓ of animal k in herd i making a record at age j.

The g's are uncorrelated, the p's are uncorrelated, the w's are uncorrelated and the g's, p's, and w's are mutually uncorrelated.

Heritability is .10.

Repeatability is .40.

Phenotypic variance is 64.

The following records are available to estimate herd and age effects.

a) Set up numerically (but do not solve) the mixed model equations corresponding to the model and the records.

There are: 2 herds, 2 ages, 5 animals

Herd	Age	Animal	i,j,k,ℓ	$y_{ijk\ell}$
1	1	1,1	1,1,1,1	114
1	2	1,1	1,2,1,2	118
1	2	1,2	1,2,2,1	108
2	2	2,1	2,2,1,1	120
2	1	2,2	2,1,2,1	118
2	2	2,2	2,2,2,2	122
2	1	2,3	2,1,3,1	116
2	2	2,3	2,2,3,2	120

b) Show how the equations would be modified if the relationship matrix and its inverse were:

$$A = \begin{array}{c} \\ 1,1 \\ 1,2 \\ 2,1 \\ 2,2 \\ 2,3 \end{array} \begin{array}{c} \begin{array}{ccccc} 1,1 & 1,2 & 2,1 & 2,2 & 2,3 \end{array} \\ \begin{pmatrix} 1 & 1/4 & 0 & 0 & 0 \\ 1/4 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array} ; \quad A^{-1} = \frac{1}{15} \times \text{each element} \begin{pmatrix} 16 & -4 & 0 & 0 & 0 \\ -4 & 16 & 0 & 0 & 0 \\ 0 & 0 & 20 & -10 & 0 \\ 0 & 0 & -10 & 20 & 0 \\ 0 & 0 & 0 & 0 & 15 \end{pmatrix}$$

4. (15 points) My last chance to stress the importance of expected values!! This is an easy problem, if you can manage to read all of it. Good luck!!

The one-way classification, fixed effects model is:

$$y_{ij} = \mu + F_i + w_{ij}$$

where y_{ij} is the record on the jth animal exposed to fixed effect i,

μ is an unknown constant,

F_i is the ith fixed effect, and

$w_{ij} = G_j + E_j$.

Note: $E(F_i) = F_i$,

$E(w_{ij}) = 0$ and

$E(w_{ij}^2) = \sigma_w^2$.

The expectation of the product of any w with any other w is zero.
The least squares equations are:

$$\begin{aligned} n_. \hat{\mu} + n_1\hat{F}_1 + n_2\hat{F}_2 + n_3\hat{F}_3 &= y_{..} \\ n_1\hat{\mu} + n_1\hat{F}_1 &= y_{1.} \\ n_2\hat{\mu} + n_2\hat{F}_2 &= y_{2.} \\ n_3\hat{\mu} + n_3\hat{F}_3 &= y_{3.} \end{aligned}$$

where $n_.$ is the total number of records,

n_i is the number of records for fixed effect i,

$y_{..}$ is the total sum of records and

$y_{i.}$ is the total sum of records for fixed effect i.

One constraint is needed to solve the equations.

a) When the constraint is that the weighted sum of fixed effects equals zero:
$n_1\hat{F}_1 + n_2\hat{F}_2 + n_3\hat{F}_3 = 0$, then $\hat{\mu} = y_{..}/n_.$

What is the expected value of $\hat{\mu}$?

b) An often used constraint is that the sum of fixed effects equals zero:
$\tilde{F}_1 + \tilde{F}_2 + \tilde{F}_3 = 0$. (~ indicates different constraint used.)

Then $\tilde{\mu} = \frac{1}{3}(\frac{y_{1.}}{n_1} + \frac{y_{2.}}{n_2} + \frac{y_{3.}}{n_3})$

What is the expected value of $\tilde{\mu}$?

4. (continued)

c) When the constraint is $\overset{*}{\mu} = 0$, what is the expected value of $\overset{*}{\mu}$?

(* indicates different constraint used.)

d) Would you expect $\hat{\mu} = \tilde{\mu} = \overset{*}{\mu}$? (That is, numerical solutions for $\hat{\mu}$, $\tilde{\mu}$, and $\overset{*}{\mu}$ to be equal.)

e) Would you expect when all the n_i are very large that $\hat{\mu} = \mu$?

Note: $\hat{\mu} = y_{..}/n_.$ and $n_. = n_1 + n_2 + n_3$.

f) What is $E(\hat{F}_1)$,

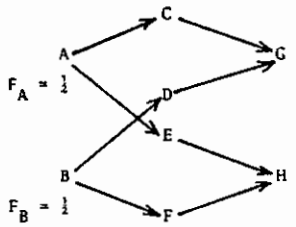
$E(\tilde{F}_1)$, and

$E(\overset{*}{F}_1)$?

FIRST EXAM, 1983

(30 points)

1. The sequence of birthdates of the animals in the following arrow diagram corresponds to the letters (A is the oldest).



The inbreeding coefficient of A is $\frac{1}{2}$ and of B is also $\frac{1}{2}$.

a) Calculate the additive relationships among the eight animals.

b) Which are inbred? What are the inbreeding coefficients?

c) If G is mated to H to produce I, what is the additive relationship of G to I?

d) What is the inbreeding coefficient of I?

e) What is the dominance relationship between D and F?

f) What is the dominance relationship between G and H?

(30 points)

2. The following are tables of

additive relationships, a_{ij}

dominance relationships, d_{ij}

environmental covariances divided by the phenotypic variance, c_{ij}

	a_{ij}					d_{ij}					c_{ij}				
	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E
A	1	0	1/2	1/2	3/8	1	0	0	0	0	-	0	1/8	1/8	0
B	0	1	1/2	1/2	0	0	1	0	0	0	0	-	0	0	0
C	1/2	1/2	1	1/2	3/16	0	0	1	1/4	0	1/8	0	-	1/4	0
D	0	1/2	1/2	1	3/16	0	0	1/4	1	0	1/8	0	1/4	-	0
E	3/8	0	3/16	3/16	5/4	0	0	0	0	1	0	0	0	0	-

The following variances are known:

$\sigma^2_x = 100$

$\sigma^2_{10} = 20$

$\sigma^2_{20} = 8$

$\sigma^2_{11} = 4$

$\sigma^2_{01} = 12$

$\sigma^2_{PE} = 18$

a) What is repeatability?

b) What is heritability in the narrow sense?

What are the covariances between records of:

c) (A and C)?

d) (C and D)?

e) The covariance between two records of E?

(30 points)

3. Set up symbolically and numerically but do not solve the equations that determine the weights to predict additive genetic value for animals A, B, and C.

	Own record		Records of progeny of animal (half sibs of each other)		Record of sire of animal	
Animal	Number	Average	Number	Average	Number	Average
A	3	50	21	30	1	-10
B	3	-20	21	60	1	40
C	3	70	21	20	1	30

Assume: Heritability (narrow sense) = .40

Repeatability = .60

Dominance variance = 20

Phenotypic variance = 200

Environmental covariances

between full sibs = 40

between half sibs = 10

(10 points)

4. Cloning can produce identical twins having completely different uterine and postnatal environments. Assume the covariance between pairs of such genetically identical clones is 90.

The covariance between pairs of identical twins raised in the same maternal environment is 120.

The covariance between pairs of paternal half sibs (have same father but unrelated mothers) raised in different environments is 15.

The phenotypic variance for the population is 200.

a) Interpret these covariances assuming that the only genetic effects are due to additive and to dominance effects, i.e., $\sigma^2_G = \sigma^2_{10} + \sigma^2_{01}$.

For example, what is the additive heritability and what is heritability in the broad sense?

b) If you have trouble starting this problem, the TA will give you a hint (but you can then earn a maximum of 8 points for this problem).

SECOND EXAM, 1983

1) (25 points)

Animals A, B, and C are rare identical triplets. Each has two records. Their averages, adjusted for the mean and other factors, are:

$X_A = 10$

$X_B = 5$

$X_C = 15$

Assume:		
	Additive genetic variance	= 100
	Total genetic variance	= 120
	Phenotypic variance	= 400
	Repeatability	= .50
	Environmental covariance	= 0

a) Predict the additive genetic value of animal A. If you are uneasy working with triplets, you can, for a penalty of only one point, change the problem to involve identical twins, A and B.

b) Predict a future record of animal A. If you were working with twins, you may continue with no further penalty.

2) (25 points)

The average of 41 daughters of a bull is +2000. With heritability of .40, the weighting factor to predict one-half of his additive genetic value is $p/(p+9)$, where p is the number of daughters. Thus, the prediction of his transmitting ability is:

$$I = (41/50)(2000) = 1640 \quad .$$

The prediction of the average additive genetic value of an infinite number of his future progeny is also 1640.

The prediction of a record of a future progeny is also 1640.

Assume: $h^2 = .40$
$r = .60$
$\sigma_x^2 = (2500)^2$

a) What is the probability the TA of the bull is greater than 0?

b) What is the 90% confidence range for the average additive genetic value of an infinite number of future progeny?

c) What is the probability the record of a future progeny is greater than 1640?

d) What is the probability the record of a future progeny is greater than 0?

Do either of (3i) or (3ii), not both.

3i) (25 points)

Given: Phenotypic variances
milk yield (2000 lb)^2
protein yield (100 lb)^2

phenotypic covariance (.8)(2000)(100) = 160,000

additive genetic variances
milk yield .25(2000 lb)^2 = (1000 lb)^2
protein yield .16(100 lb)^2 = (40 lb)^2

additive genetic covariance
(.7)(1000)(40) = 28,000

The covariance between the milk and protein averages for p daughters of the same bull is

$$\frac{\text{phenotypic cov.} + (p-1)(.25 \text{ genetic cov.})}{p}$$

Assume a bull has 50 daughters, each with a milk and a protein yield record. Let X_M = average of the 50 milk records and X_P = average of the 50 protein records.

Compare two selection indexes to predict additive genetic value for milk:

$I_1 = bX_M$ $\quad b = 1.5385$

and

$I_2 = b_M X_M + b_P X_P$ $\quad b_M = 1.7014$, $b_P = -5.2634$

a) What are the accuracy values for the two indexes for predicting additive genetic value for milk yield?

b) If selection is on I_1, what is the expected correlated response for protein yield?

c) If selection is on I_2, what is the expected correlated response for protein yield?

Do either of (3i) or (3ii), not both.

3ii) (25 points)

Given: Phenotypic variances
milk yield (2000 lb)^2
protein yield (100 lb)^2

phenotypic covariance (.8)(2000)(100) = 160,000

additive genetic variances
milk yield .25(2000 lb)^2 = (1000 lb)^2
protein yield .16(100 lb)^2 = (40 lb)^2

additive genetic covariance
(.7)(1000)(40) = 28,000

Assume a cow has both a milk record and a protein record, X_M and X_P.

Compare two selection indexes for predicting additive genetic value for milk yield:

$$I_1 = bX_M$$

and

$$I_2 = b_M X_M + b_P X_P \quad .$$

a) What are the accuracy values for the two indexes for predicting additive genetic value for milk yield?

b) If selection is on I_1, what is the expected correlated response for protein yield?

c) If selection is on I_2, what is the expected correlated response for protein yield?

4) (25 points)

A friend asks you to recommend the better of two breeding programs for improving additive genetic value that your friend is considering. Show, numerically, the relative expected gain per year for the two plans.

Plan I. A) Select sires based only on records of their mothers.
Assume: 1) The r_{TI} for the additive genetic value of mothers is .60,
2) Selection is from the best 1 of 50, and
3) Generation interval is 2 years for sires.

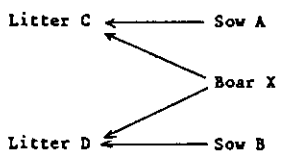
B) Select dams based only on their records.
Assume: 1) The r_{TI} for dams is .60,
2) Selection is for the best 90%, and
3) Generation interval is 4 years for dams.

Plan II. A) Select sires based on records of 50 daughters (an AI stud maintains and tests the sires).
Assume: 1) r_{TI} for additive genetic value of the sires is .90,
2) Selection is for the best 1 of 10, and
3) Generation interval is 7 years for sires.

B) Select dams based on their records and evaluations of their sires.
Assume: 1) The r_{TI} for dams is .65,
2) Selection is for the best 90%, and
3) Generation interval is 4 years for dams.

FINAL EXAM, 1983

1. (25 points) (This problem should be easy.)



Record of sow A: +10

Record of sow B: -5

Average for the 6 pigs in litter C: +2

Average for the 10 pigs in litter D: +4

Phenotypic variance = 100

Additive genetic variance = 20

Dominance genetic variance = 8

Additive by additive genetic variance = 4

Environmental covariance between full sibs = 10

Environmental covariance between half sibs = 0

Environmental covariance between progeny and dam = 5

Set up numerically (but do <u>not</u> solve) the selection index equations to:

a) predict the additive genetic value of boar X.

b) predict the additive genetic value of sow A.

Then show for both cases how you would use the weights (symbolic) with the records (numeric) to:

c) predict the additive genetic value of boar X.

d) predict the additive genetic value of sow A.

2. (25 points) (This problem may seem difficult; if you have difficulty, start with $T = G_A$ and then expand to $T = G_A + G_D$.)

Additive genetic value (G_A) and dominance genetic value (G_D) for the same trait can be considered to be separate traits although the phenotypic observations will include both kinds of effects.

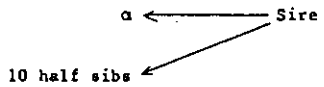
Assume selection is for $T = G_A + G_D$; note that the economic values are equal.

Assume the selection intensity factor for both selection plans I and II is $D = 1.5$.

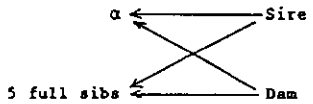
Assume $\sigma^2_{10} = 40$, $\sigma^2_{01} = 20$, $\sigma^2_p = 200$ apply to some carcass characteristic which precludes measurement on breeding animals.

Compare two selection plans.

I. Select males based on the average of single records of 10 of their half sibs.



II. Select males based on the average of single records of 5 of their full sibs.



For both plans:

a) Determine the expected superiorities in T of selected males.

b) Determine the expected superiorities in G_A of selected males.

c) Determine the expected superiorities in G_D of selected males.

d) If the selected animals are mated to a random and unrelated group of females, what fractions of the expected superiorities will be transmitted to their progeny?

3. (25 points) (This problem is not difficult except possibly for d); the easiest part is c).)

$X_1 = 1600$ is the average of single records of 45 progeny of Sire 1 by unrelated dams.

$X_2 = 1200$ is the average of single records of 25 progeny of Sire 2 by unrelated dams.

Sire 1 and Sire 2 have the same father but are not related in any other way.

Heritability = .40, phenotypic variance = $(2000)^2$.

The prediction of the additive genetic value of Sire 1 is:

$$I_{G_1} = 1.6534X_1 + .0637X_2 = 2722$$

The prediction of the additive genetic value of Sire 2 is:

$$I_{G_2} = .1147X_1 + 1.4495X_2 = 1923$$

The prediction of the additive genetic value of Sire 1 minus the additive genetic value of Sire 2 is:

$$I_{G_1-G_2} = 1.5387X_1 - 1.3858X_2 = 799$$

a) What is the probability the additive genetic value of Sire 1 is greater than 2000?

b) What is the 80% confidence range for the additive genetic value of Sire 1?

c) If only progeny of Sire 1 had been used to predict the additive genetic value of Sire 1, what would be the 80% confidence range?

d) What is the probability the difference in additive genetic values (Sire 1 - Sire 2) is greater than 500?

FIRST EXAM, 1984

1. A friend has determined that an appropriate index should be:

$$I = 2X_1 + 3X_2 .$$

The friend also supplies the following variances and covariances:

$\sigma^2_{X_1} = 10$ $\quad \sigma_{X_1X_2} = 5$ $\quad \sigma_{X_1T} = 4$ $\quad \sigma^2_T = 8$

$\sigma^2_{X_2} = 20$ $\quad \sigma_{X_2T} = 6$

a) What is r^2_{TI} for the index $I = 2X_1 + 3X_2$?

Another friend insists the correct index is

$$I = (40\ X_1 + 60\ X_2)/100$$

b) What is r^2_{TI} for the index $I = (40\ X_1 + 60\ X_2)/100$?

c) Are either of the indexes "the selection index"?

2. With patience, perserverance, persuasion, etc... a molecular biologist for a unique (imaginary) species with 6 loci has determined the identity by origin of the genes carried by animals X and Y to be as follows:

Animal X	Animal Y
a_1a_2	a_1a_3
b_2b_3	b_2b_3
c_4c_7	c_2c_5
d_5d_5	d_5d_5
e_1e_8	e_1e_8
f_4f_4	f_3f_4

a) What is the inbreeding coefficient of animal X?

b) What is the additive relationship between X and Y?

c) What is the dominance relationship between X and Y?

3. In introductory animal genetics courses, a formula is given for predicting the breeding value (additive genetic value) of an animal from the average of its records. You are now able to derive that formula. A formula usually is also given for predicting the real producing ability of an animal from the average of its records. You are now able to derive that formula.

Animal i has 4 records averaging +6.

Animal j has 2 records averaging +8.

$h^2 = .30$ $r = .60$ $\sigma_X^2 = 81$

a) Predict the breeding value of i and of j.

b) Predict the real producing ability of i and j.

c) Can real producing ability be predicted directly from estimated breeding value without estimating real producing ability from the formula used in b)? Why?

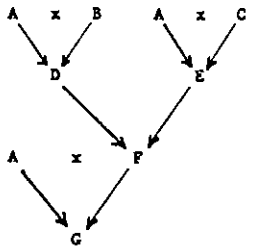
For d), e), and f), assume animal i and animal j are both inbred with $F_i = F_j = .25$.

d) Repeat a).

e) Repeat b).

f) Repeat c).

4. A sheep breeder bred the "best ram in the land" to the "best ewe in the land." The mating resulted in triplets (A, B, C) that looked so much alike that the breeder thought they were identical triplets. They also looked so much like the "best in the land" parents that the breeder decided to establish a line of "best in the land" sheep based on the triplets using the breeding plan described in the arrow diagram. Blood tests later determined B and C were identical females; A was a male!!



a) Show additive relationships among animals A through G on the next blank page.

b) Which animals are inbred and what are their inbreeding coefficients?

c) What is the additive relationship of G to the "best ram in the land"?

For d) and e), assume $\sigma_{10}^2 = 16$, $\sigma_{01}^2 = 8$, $\sigma_{20}^2 = 32$ and $\sigma_E^2 = 44$.

d) What is the covariance between a record of A and a record of B?

e) What is the covariance between a record of B and a record of C?

SECOND EXAM, 1984

1. This is a single trait question. (30 points)

Assume:

$\sigma^2_{10} = 20$

$\sigma^2_{01} = 16$

$\sigma^2_{20} = 8$

$\sigma^2_{PE} = 16$

$\sigma^2_{X} = 100$

The environmental covariance between full sibs is .04(100) = 4.

a) A record of animal 1 is $X_1 = 8$.

A record of animal 2 is $X_2 = 12$.

Animal 1 and animal 2 are full sibs.

i) Predict the additive genetic value of animal 1.

ii) What is the r_{TI}?

b) The average of single records of animal 1 and animal 2 is $X_1 = 10$. They are full sibs in the same litter. Only the litter average can be measured.

i) Predict the additive genetic value of animal 1.

ii) What is the r_{TI}?

iii) Predict the additive genetic value of a full sib of animals 1 and 2.

iv) What is the r_{TI}?

v) Predict the average genetic value of animals 1 and 2.

vi) What it the r_{TI}?

2. This is a single trait question. (30 points)

Given: Phenotypic variance = 200

Repeatability = .50

Heritability = .25

Average of n = 3 records of animal Z = 20

T = real producing ability of Z

I = [n/(n+1)](20) = 15

$r_{TI} = \sqrt{n/(n+1)} = \sqrt{.75}$

a) What is the probability that the real producing ability of Z is greater than 15?

b) What is the probability that the real producing ability of Z is greater than 10?

c) What is the probability that the next record of Z is greater than 15?

d) What is the probability that the next record of Z is greater than 10?

3. For two traits, 1 and 2, the following is known (40 points)

$\sigma^2_{X_1} = 100$ $h^2_1 = .25$

$\sigma^2_{X_2} = 400$ $h^2_2 = .36$

$\sigma_{X_1X_2} = 150$ $r_{g1g2} = .80$

Measurements on traits 1 and 2 for animals A and B are:

	X_1	X_2
A	10	-20
B	-10	20

Animals A and B are unrelated.

Assume the net economic values are:

$5 per unit of trait 1

$10 per unit of trait 2.

a) Predict overall genetic value of animals A and B.

b) What are the relative economic values of traits 1 and 2?

Assume the correct index is: $-.5X_1 + 4X_2$.

Assume selection is based on this index and the largest 1 of 50 is selected.

c) What is the expected superiority in overall genetic value of the selected 1 of 50?

d) What is the expected correlated response in trait 1?

(25 points)

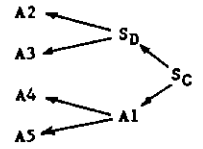
1. This is the Mixed Model Equations problem.

Given:

Animal	Age of Dam (Yr)	Sex	Weaning Weight of animal
A1	2	Bull = 1	500
A2	2	Heifer = 2	300
A3	3	2	400
A4	3	2	200
A5	3	1	600

Age of dam is known to affect weaning weight.
Sex of animal is known to affect weaning weight.
Additive genetic value of the animal is known to affect weaning weight.

Heritability of weaning weight is .20, $\sigma^2_y = 400$. Animals 1, 2, 3, 4, 5 are related through bulls C and D as described in the arrow diagram:



Set up the mixed model equations to predict the additive genetic values of A1, A2, A3, A4 and A5. (Also write the model you are using.)

(25 points)

2. This is the selection index problem for a single trait. Read carefully.
Part (a) will be to predict the additive genetic value of animal C.
Part (b) will be to predict the additive by additive genetic value of animal C.

Given:

The sire of C is animal Z (no records).
The dam of C is animal D.
The average of 4 records of animal D is $X_2 = +12$.
Animals A and B are full sibs.
The parents of A and B are animals S and D.
The average of single records of animals A and B is $X_1 = +8$.
The inbreeding coefficient of animal C is $F_C = .25$.

Assume:

$\sigma^2_{10} = 20$

$\sigma^2_{01} = 16$

$\sigma^2_{20} = 8$

$\sigma^2_{PE} = 16$

$\sigma^2_X = 100$

The environmental covariance between full sibs is .04(100) = 4.

(20 points)

a) Predict the additive genetic value of animal C.

(5 points)

b) Predict the additive by additive genetic value of animal C.

(25 points)

3. This is the probability problem.

Given: $X_S = 24$ is a record of animal S

$X_T = -9$ is a record of animal T

S and T are half-sibs, $a_{ST} = 1/4$

$h^2 = .40$, $r = .60$, $\sigma^2_p = 100$

(10 points)

a) To predict the additive genetic value of S, G_{A_S}, the S.I. equations are:

$$100b_S + 10b_T = 40$$

$$10b_S + 100b_T = 10$$

$$I = (13/33)(24)+(2/33)(-9) = 8.9$$

What is the probability the additive genetic value of animal S is greater than 14?

(5 points)

b) Predict the additive genetic value of animal T?

(10 points)

c) To predict the difference in additive genetic values between animals S and T, $G_{A_S} - G_{A_T}$, the S.I. equations are:

$$100b_S + 10b_T = 30$$

$$10b_S + 100b_T = -30$$

$$I = (1/3)(24) - (1/3)(-9) = 11$$

What is the probability the difference in additive genetic values between animals S and T is greater than 14?

(25 points)

4. This is the multiple trait and genetic superiority problem. (All genetic values are assumed to be additive genetic values.)

Given:

$G_{1\alpha}$ is genetic value for trait 1.
$G_{2\alpha}$ is genetic value for trait 2.

Overall genetic value of animal α = $T_\alpha = 5G_{1\alpha} + 2G_{2\alpha}$

$X_{1\alpha}$ = 40 (average of 5 records on trait 1 for animal α)
No records on trait 2.

Trait	Heritability	Repeatability	Phenotypic standard deviation
1	.36	.50	10
2	.16	.25	20

genetic correlation = -.50
phenotypic correlation = .20

(10 points)

a) Predict overall genetic value for animal α.

(2 points)

b) Will changing the economic value of trait 2 affect ranking of animals? Explain!!

4. continued on next page.

4. (Continued.) For parts c, d, and e:

Assume selection for overall genetic value is based on $I^* = X_1$ (average of 5 records for the animal being evaluated.)
Assume 2 animals out of 5 are selected.

(5 points)

c) What is expected genetic superiority for trait 2? Comment??

(3 points)

d) What is expected genetic superiority for trait 1?

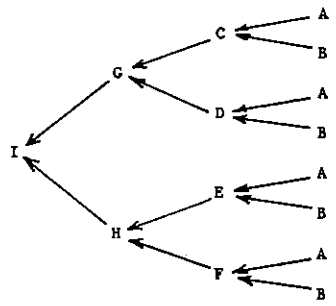
(5 points)

e) What is expected genetic superiority for $T = 5G_1 + 2G_2$?

First Exam October, 1985

1. (25 points)

Given: $F_A = 1$, $F_B = 1$, $a_{AB} = 0$ and the following pedigree



a) Compute the table of additive relationships.

b) What is the <u>additive relationship</u> between A and I?

c) Between G and H?

d) What is the <u>coefficient of relationship</u> between G and H?

e) What is the <u>inbreeding coefficient</u> of I?

f) What is the <u>dominance relationship</u> between C and E?

2. (25 points)

Let $\sigma^2_{10} = 64$, $\sigma^2_{01} = 32$, $\sigma^2_{20} = 16$, $\sigma^2_{PE} = 38$, $\sigma^2_{TE} = 50$

The following are tables of a_{ij}, d_{ij}, and c_{ij} for animals A, B, C, D, E.

"a"

	A	B	C	D	E
A	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	0
B		1	$\frac{1}{2}$	$\frac{1}{4}$	0
C			1	$\frac{1}{4}$	0
D	Symmetric			1	0
E					$1\frac{1}{2}$

"d"

	A	B	C	D	E
A	1	$\frac{1}{4}$	$\frac{1}{4}$	0	0
B		1	$\frac{1}{4}$	0	0
C			1	0	0
D	Symmetric			1	0
E					1

"c"

	A	B	C	D	E
A	-	.10	0	.05	.20
B		-	0	.05	.20
C			-	0	.02
D	Symmetric			-	.05
E					-

Calculate numerical values for:

a) Heritability.

b) Repeatability.

c) Phenotypic variance for records of animals such as A.

d) Phenotypic variance for records of animals such as E.

e) Covariance beween a record of A and a record of B.

f) Covariance between a record of A and a record of D.

g) Covariance between a record of A and a record of E.

h) Covariance between a record of B and a record of C.

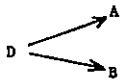
i) Covariance between two records of A.

j) Covariance between two records of E.

3. (25 points)

Let $h^2 = .20$ and $\sigma_X^2 = 100$

Animal D is the parent of animals A and B that are otherwise unrelated.



The index to predict additive genetic value of D from

X_A, a record of A and

X_B, a record of B

is $I_D = b_A X_A + b_B X_B$

a) Write both symbolically and numerically , but do not solve, the equations to find b_A and b_B.

b) Now suppose that additive genetic value of A is to be predicted from $I_A = b_A X_A + b_B X_B$. Write both symbolically and numerically, but do not solve, the equations to find b_A and b_B.

c) Will $b_A = b_B$? Why or why not.

d) Write both symbolically and numerically, but do not solve, the equation to find b_{AB} for predicting the additive genetic value of A as $I_A = b_{AB} X_{AB}$ if only the average of records of A and B is known [that is, $X_{AB} = (X_A + X_B)/2$].

4. (25 points)

Given: $\sigma_{X_1}^2 = 100$ $\sigma_{X_1X_2} = 20$ $\sigma_{X_1T} = 40$

$\sigma_{X_2}^2 = 100$ $\sigma_{X_2T} = 20$

$\sigma_T^2 = 40$

Two proposed indexes that use X_1 and X_2 are

$I_A = X_1 + .5X_2$ and

$I_B = 12X_1 + 6X_2$

a) What is the variance of I_B?

b) What is the covariance between I_B and T?

c) What is the covariance between I_A and I_B?

d) Are either of these the "best" selection index? If yes, which one?

Second Exam November, 1985

(25 points)

1. Compare plans A and B for gain in additive genetic value. Assume $h^2 = .25$, $r = .50$ and the following:

Plan	Selection path	r_{TI}	Number Selected	Number Available	Generation interval (yr)
A	Sires of sons	.90	2	50	10
	Dams of sons	.60	150	15,000	6.5
	Sires of daughters	.85	15	50	8
	Dams of daughters	.55	12,000	15,000	5.5
B	Sires (1 record of dam)	.25	15	15,000	2
	Dams (1 record)	.50	12,000	15,000	4

(25 points)

2. Let $X_1 = 8$ = average of single records of 50 paternal half sibs of animal D.
$X_2 = 12$ = average of 3 records of the dam of D.

Assume $h^2 = .40$, $r = .60$, $\sigma_P^2 = 100$

The selection index equations for predicting additive genetic value of D are:

$$11.8b_1 + 0b_2 = 10$$

$$0b_1 + 73.\bar{3}b_2 = 20$$

Thus $I = .85(8) + .27(12) = 10.0$

(6 pts) a) What is the probability the <u>additive genetic value</u> of D <u>exceeds</u> 10?

(7 pts) b) What is the probability the <u>additive genetic value</u> of D <u>exceeds</u> 15?

(3 pts) c) What is the probability a single <u>record</u> of D, X_D, would be <u>less than</u> 0?

(3 pts) d) What is the probability a single <u>record</u> of D, X_D, would be <u>greater than</u> 25?

(3 pts) e) What is the probability the <u>real producing ability</u> of D is <u>less than</u> 10?

(3 pts) f) What is the probability the <u>real producing ability</u> of D is <u>less than</u> 15?

(25 points)

3. The following two tables contain the phenotypic and genetic variances and covariances among traits A, B, C, and D. If you do not understand the table, be sure to ask the TA to explain it.

	Phenotypic					Genetic			
	A	B	C	D		A	B	C	D
A:	100	100	120	0	A:	25	40	60	-100
B:		400	100	0	B:		10	50	-20
C:			900	0	C:			400	-150
D:				1600	D:				900

Three breeders want to improve the genetic value of trait A.

Breeder 1 is using

$$I_1 = .1X_B$$

Breeder 2 is using

$$I_2 = .1X_C - .1X_D$$

Breeder 3 is using

$$I_3 = X_A$$

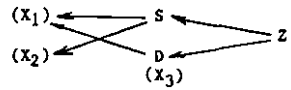
Assume equal selection intensities and generation intervals for the 3 plans:

(15 pts) a) Compare expected genetic gains for trait A for selection based on I_1, I_2, and I_3.

(10 pts) b) If $v_A = 10$ and $v_C = 5$, i.e., $T = v_A G_A + v_C G_C$, find the best index if X_A and X_B are available, i.e., $I = \beta_A X_A + \beta_B X_B$. Set up numerically but do not solve the equations.

(25 points)

4.



X_1 = average of single records of 5 full sibs (parents are S and D).

X_2 = average of two records on each of 20 paternal half sibs (parents are S and 20 unrelated dams).

X_3 = average of 4 records on D.

Note that S and D are half sibs (see diagram).

Given: $h^2 = .30$ $\sigma^2_{20}/\sigma^2_X = .10$ $r = .70$ $\sigma^2_X = 100$

$c_{11'} = .10$ $c_{12} = .05$ $c_{13} = .20$

$c_{22'} = .05$ $c_{23} = .00$

a) Set up numerically the selection index equations to predict the additive genetic value of S. Do not solve the equations. Note 1: Compute the relationship table first. Note 2: There are no formulas in the book for exactly this case.

Graduate students only!!

b) Set up numerically the selection index equations to predict a record of a full sib of the animals with records in X_1 (that full sib was born at the same time but does not have a record). Do not solve the equations.

Final Exam December, 1985

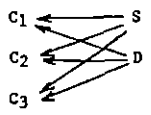
(25 points) The only problem with mixed model equations.

1. Weaning weight can be considered as a trait of the calf or as a trait of the mother. In the second case the trait can have repeated records. Age of the mother and sex of calf are known to influence weaning weight. As a trait of the mother, assume heritability is .20 and repeatability is .60 with phenotypic standard deviation of 50 lb.

   As an example of the mixed model equations, set up the equations to estimate the effects of age and sex and to predict additive genetic value and real producing ability for cows 1, 2, and 3 from the following records.

Weaning weight	Cow	Age of cow	Sex of calf
400	1	Young	Male
500	1	Old	Female
400	2	Young	Female
400	3	Young	Female
600	3	Old	Male
500	3	Old	Female

The cows are full sibs as shown:



Some hints:

How many age equations?
How many sex equations?
How many permanent environment equations?
How many additive genetic value equations?
Can you predict real producing ability directly? If not, how?

The next page is blank and waiting for the appropriate MME.

(25 points) Multiple trait problem

2. The phenotypic and genetic correlations among, and phenotypic standard deviations and heritabilities for, traits A, B and C are as follows:

	Correlations: Phenotypic A	Phenotypic B	Phenotypic C	Genetic A	Genetic B	Genetic C	Phenotypic standard deviation	Heritability
A	--	.80	-.10	--	.60	.00	50	.36
B		--	-.20		--	-.10	40	.25
C			--			--	10	.16

A breeder has been selecting for trait A with $I_A = .36X_A$. After several years genetic response for trait A was estimated to be $\hat{\Delta G}_A = 100$.

a) What is the expected correlated genetic change in trait B?

b) What is the expected correlated genetic response in trait C?

c) Traits B and C cannot be measured until late in life. Trait B can be measured on the sire and trait C can be measured on the dam. Assume the economic values for the additive genetic values are $v_A = 4$, $v_B = 2$, $v_C = 0$. Set up (but do not solve) the equations to find the selection index weights for the index

$$I = \beta_1 X_A + \beta_2 X_{B_S} + \beta_3 X_{C_D}$$

where X_A is a record on trait A of the animal being evaluated,

X_{B_S} is a record on trait B of the sire, and

X_{C_D} is a record on trait C of the dam.

Assume the sire and dam are unrelated.

The next page is blank.

(25 points)

3. Selection index problem

a) Rank the following five animals for additive genetic value. Assume heritability is .40, repeatability is .80, and phenotypic variance is 200.

Animal	Information (all as differences from the "mean")
1	Record on animal 1 = +20.
2	Record on sire of 2 = +30. Animal 1 is the dam of 2.
3	The dam of animal 3 is animal 1.
4	Average of single records on 10 progeny of animal 4 = +19 (Other parents are unrelated.)
5	Average of single records of 10 half sibs of animal 5 = +19. (The half sibs are equally related.)

b) Rank animals 2, 3, 4, and 5 for dominance genetic value. All information needed is given. Think!!

(25 points)

4. Another selection index problem

Predict the difference in genetic value (a) between animals X and Z and (b) between animals Y and Z, i.e., $G_X - G_Z$ and $G_Y - G_Z$. Animals Y and Z are full sibs and are unrelated to animal X.

Assume:

Animal X has 2 records that average +10.

Animal Y has 1 record of +8.

Animal Z has 1 record of +6.

$h^2 = .60$ $\qquad$ $c_{XY} = .00$

$r = .80$ $\qquad$ $c_{XZ} = .00$

$\sigma^2_{Dl}/\sigma^2_P = .20$ $\qquad$ $c_{YZ} = .15$

$\sigma^2_P = 100$

Predict a) $G_X - G_Z$ and b) $G_Y - G_Z$.

Go to next page.

4. (continued)

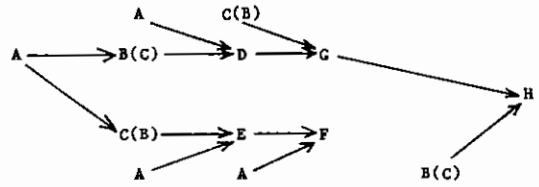
c) Compare the accuracy of prediction of $G_X - G_Z$ (difference between unrelated animals) with the accuracy of prediction of $G_Y - G_Z$ (difference between full sibs). If you were unable to solve a) and b) show how you would do the calculations.

d) How will the equations change if the difference in future records of animals Y and Z is to be predicted?

First Exam October, 1986

(25 points)

1. In the following pedigree, arrows go from parent to progeny. Animals B(C) and C(B) are the result of splitting an embryo and thus are related as identical twins.

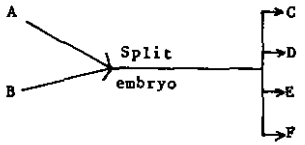


a) Calculate the additive relationships among the eight animals.

b) Which animals are inbred and what are their inbreeding coefficients?

(25 points)

2. Animals A and B are mated. The fertilized egg is split so that four identical clones are born: C, D, E, and F.



Let the average of five adjusted records of A be $X_1 = 40$ and the average of adjusted single records of C, D, E, F be $X_2 = 30$.

Assume $\sigma^2_{10} = 50$ $\sigma^2_{01} = 20$ $\sigma^2_{20} = 30$ $\sigma^2_X = 200$

$h^2 = .25$ $r = .50$.

Set up symbolically and numerically the equations to find the weights for X_1 and X_2 to use to predict the genetic value of animal B.

(50 points:equal value to all parts)

3. Animals A and B have records X_1, X_2, and X_3 available for predicting true value, T. Animals C and D have X_1 and X_2. Assume $\sigma^2_T = 40$.

Animal	X_1	X_2	X_3
A	20	40	10
B	0	100	50
C	40	20	Missing
D	20	40	Missing

The selection index to predict T for A and B is:

$$I = .40X_1 + .10X_2 + .05X_3.$$

The selection index equations to determine the weights are:

$$\begin{aligned} 100b_1 + 0b_2 + 10b_3 &= 40.5 \\ 0b_1 + 200b_2 + 20b_3 &= 21.0 \\ 10b_1 + 20b_2 + 80b_3 &= 10.0 \end{aligned}$$

a) Rank animals A and B for T.

b) Calculate r^2_{TI} for this index.

c) Calculate σ^2_I for this index.

d) What is σ_{TI} for this index?

e) For an animal with I = 12.5, what is $V(T|I = 12.5)$?

f) For an animal with I = 2.5, what is $V(T|I = 2.5)$?

g) Can animals C and D be ranked if X_3 is missing? The answer is yes!! Write yes for full credit.

h) What is the selection index to be used for animals C and D?

i) Rank animals A, B, C and D for T.

j) Which index, the one for A and B or the one for C and D, has the larger r_{TI}? (A numerical answer is best; a verbal answer for part credit.)

k) Which index has the larger σ_I^2? (A numerical answer is best; a verbal answer for part credit.)

l) Which index is associated with the larger $V(T|I = I_0)$? (A numerical answer is best; a verbal answer for part credit.)

Another index has been proposed:

$$I = .2X_1 + .1X_3.$$

m) Will this index have a larger r_{TI} than $I = .40X_1 + .10X_2 + .05X_3$? (A numerical answer is best; a verbal answer for part credit.)

n) Calculate $V[T - (.2X_1 + .1X_3)]$.

Second Exam November, 1986

1. (25 points)

More split embryo, identical twins!!

The additive relationship between full sib parents of identical twins W and Z is .50.

Twin W has 5 records (adjusted for μ) averaging, $X_W = -10$.

Twin Z has 4 records (adjusted for μ) averaging, $X_Z = +12$.

Given: $\sigma_{IO}^2 = 40$, $r = .60$, $\sigma_X^2 = 100$

a) Set up symbolically and numerically (but do not solve) the equation(s) to find the weight(s) to predict the additive genetic value of twin W.

b) Will the predicted additive genetic value of twin Z be the same as for twin W? Why or why not?

c) Set up symbolically and numerically (but do not solve) the equation(s) to find the weight(s) to predict a future record of twin W.

d) Will the predicted future record of twin Z be the same as for twin W? Why or why not?

2. (25 points)

A group of breeders with small herds asked their extension agent why genetic progress in their herds seems to be less than in a group of similar herds where all animals are pooled for selection. The selection methods are exactly the same for both sets of breeders. Selection of males is based on a performance record. There is no selection on females. The generation interval for males is 2 years and for females is 3 years. Heritability of the trait is .36 and the additive genetic standard deviation is 600.

In the small herds, in each generation each breeder has available only 5 males from which the top one is selected based on his record.

In the pooled herds, 10,000 males are available from which 20% are chosen based on their records and randomly assigned to the cooperating herds.

a) Do you have an explanation?

b) Calculate expected genetic gain.

3. (25 points)

Relative economic values (values per phenotypic standard deviation) for traits A and B are 1,000 and -4,000, respectively. Phenotypic and genetic variances of and covariances between the traits are: (If you are not sure what the tables mean, see the TA).

	Genetic		Phenotypic	
	A	B	A	B
A	50	20	100	50
B	20	40	50	400

a) A breeder measures traits A and B on all animals. Set up symbolically and numerically (but do not solve) the equations to find the selection index. For the correct I, $\sigma_I = 350$.

b) The breeder tells you that trait B is measured late in life and extends the generation interval an extra 3 years: from 3 years to 6 years. The breeder asks you if selection is on trait A alone, what will the economic gain be as compared to waiting for trait B and selecting using both traits A and B? Assume the selection intensity factor is 1.00 for selection, both on $I = \beta_A X_A + \beta_B X_B$ and on $I = 10X_A$ (this is the correct selection index if only X_A is used; if you prefer you can use $I = X_A$ because the ranking will be the same; for $I = 10X_A$, $\sigma_I = 100$.

c) You have what you think is a good idea and suggest to the breeder the possibility of using the dam's record for trait B instead of the animal's record for trait B which would not extend the generation interval of 3 years. Assume the dam's record for trait B will be available at the time of selection. The selection index would be:

$$I = 10.8X_A - 7.8X_{B(Dam)}$$

with $\sigma_I = 185$. Compare your idea with that of the breeder.

d) Compute expected responses for G_A and G_B for the index in c),

$$I = 10.8X_A - 7.8X_{B(Dam)}$$

4. (25 points, maximum loss is 4 points for part c).

A dairy manager has recently been told to minimize risk in choosing bulls. Risk is rather difficult to define in terms I can understand. The following problem has something to do with risk.

Transmitting ability has been predicted for ten unrelated bulls, each having 35 daughters. Assume for milk yield that heritability is .25, repeatability is .50, additive genetic standard deviation is 1,000 lb and phenotypic standard deviation is 2,000 lb. Note that b = .7.

The indexes for transmitting ability are ranked as follows:

Bull:	1	2	3	4	5	6	7	8	9	10
ETA:	1000	900	800	700	600	500	400	300	200	100

a) What is the probability that the transmitting ability of Bull 1 is less than 500?

b) What is the probability that a record of a future daughter of bull 1 will be less than 500? Note that b = .7.

c) To minimize risk, the general recommendation is to use a group of bulls rather than a single bull. The dairy manager decides to use bulls 1, 2, 3, 4, and 5 equally.

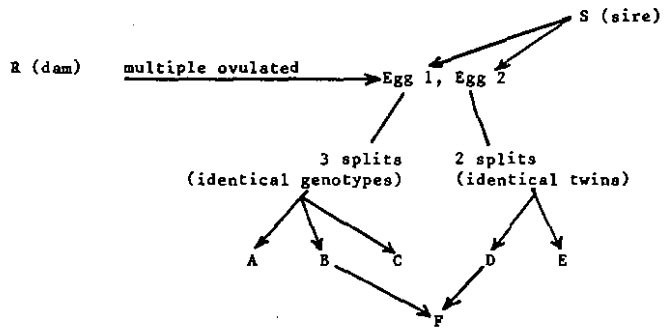
What is the probability that the average transmitting ability of the 5 bulls is less than 500? Note that

$$(1-r_{TI}^2)\sigma_T^2 = \sigma_T^2 - r_{TI}^2\sigma_T^2 = \sigma_T^2 - \sigma_I^2$$

(OMIT) d) What is the probability if each of the 5 bulls has 1 future daughter that the 5 daughters will average less than +500 as compared to daughters of bulls with ETA = 0?

Final Exam December, 1986

1. (25 points) More identical twins and split embryos!!



Given: 100 = additive genetic variance

60 = additive by additive genetic variance

40 = dominance genetic variance

50 = environmental covariance between members of the same clone

20 = environmental covariance between full sibs from multiple ovulation

.50 = repeatability

500 = phenotypic variance

The average of single records of A, B, and C = $(X_A + X_B + X_C)/3 = 150 = X_1$.

The average of single records of D and E = $(X_D + X_E)/2 = 100 = X_2$.

a) Set up (but do not solve) the equations to find the weights to predict the additive genetic value of animal F. The parents of F are B and D.

b) What is the numerical value of σ^2_I for the evaluation in a)?

2. (25 points)

Given: $\sigma^2_{I0} = 100$ $\sigma^2_{20} = 40$ $\sigma^2_X = 400$

$h^2 = .25$ $r = .50$

X_A = average of 5 records on animal A = +24

X_B = average of 2 records on animal B = +15

A and B are unrelated.

To predict the genetic value of animal A:

$$I_{A,G} = b_{A,G}X_A = (7/12)(24) = 14$$

To predict the genetic value of animal B:

$$I_{B,G} = b_{B,G}X_B = (7/15)(15) = 7$$

To predict real producing ability of animal A:

$$I_{A,R} = b_{A,R}X_A = (5/6)(24) = 20$$

To predict real producing ability of animal B:

$$I_{B,R} = b_{B,R}X_B = (2/3)(15) = 10$$

To predict the difference in future records of A and B:

$$I_{A-B,F} = b_A X_A + b_B X_B = (5/6)(24) - (2/3)(15) = 10$$

a) What is the probability that the genetic value of animal B is greater than 0?

b) What is the probability that a future record of animal B will be greater than 20?

c) What is the probability that a future record of animal A will be greater than a future record of animal B?

3. (25 points)

Given: X_1 = single record on trait A for animal J.

X_2 = single record on trait B for animal J.

X_3 = single record on trait A for animal K.

Net economic values of traits A, B, and C are 0, 0, and \$10, per unit of the three traits, respectively.

Animals J and K are full sibs of different sexes.

Variances and covariances

	Genetic				Phenotypic		
	A	B	C		A	B	C
A:	100	10	60	A:	500	40	60
B:	10	200	-50	B:	40	400	-50
C:	60	-50	150	C:	60	-50	600

a) Set up (but do not solve) the equations to predict the economic genetic value of animal K.

b) What are the relative economic values?

Continued on next page

3. Continued

Another breeder assigns net economic values of \$.50, \$2.00, and \$6.00 to traits A, B, and C. For those economic values the index for animal K is $I_K = .341X_1 + .087X_2 + .825X_3$.

c) What are the relative economic values?

d) Assume $I_K = 100$. What is the 68% symmetrical confidence range for $T_K = \$.50(G_A) + \$2.00(G_B) + \$6.00(G_C)$ given $I_K = \$100$?

e) Calculate the expected response in overall economic value for selection of the best 14%.

f) Calculate the expected correlated genetic response in trait C for selection of the best 14%.

4. (25 points)

S_Y is the sire of bulls S_A and S_B.

S_X is the maternal grandsire of bulls S_A and S_B and is the sire of S_C.

S_A, S_B, and S_C have progeny with weaning weight records.

The age of the dam of each progeny is known and is thought to affect weaning weight.

The sex of each progeny is known and is thought to affect weaning weight (1 = female, 2 = male).

Heritability is .20.

Phenotypic variance is $(50\ lb)^2$.

For this illustrative problem, assume only the following records are available:

Calf	Sire	Age of dam (yr)	Sex	Weaning weight
1	A	2	1	400
2	A	2	2	440
3	A	3	2	500
4	B	3	1	550
5	B	3	1	510
6	C	2	2	450
7	C	2	2	420
8	C	3	1	460

a) Write the model for a weaning weight record.

Continued on next page

b) Set up but do not solve the mixed model equations to predict the transmitting abilities of the bulls.

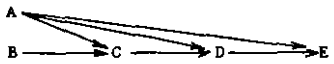
c) The European market wants sire evaluations in terms of estimated breeding values. How can you provide EBV's?

First Exam October, 1987

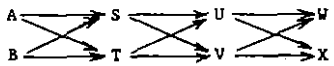
(25 points)

1. Two regular systems of mating to increase inbreeding rapidly are a) matings of a sire to his daughters, granddaughters, etc., and b) matings to full sibs. These systems are shown below.

a)



b)



For a) i) Calculate the inbreeding coefficients of C, D, and E:

ii) Is there a regular pattern of increase in the inbreeding coefficient?

iii) If A is mated to E to produce G, what will be the inbreeding coefficient of G?

For b) iv) Calculate the inbreeding coefficients of S, U, and W.

v) Is there a regular pattern of increase in the inbreeding coefficient?

vi) If W is mated to X to produce Y, what will be the inbreeding coefficient of Y?

(25 points)

2a. Assume Tom Short has calculated the covariance between paternal half-sibs (same father, different mothers) to be 1000.

Assume Kevin Wade has calculated the covariance between maternal half-sibs (same mother, different fathers) to be 1500.

Assume the calculations are the population covariances between paternal half-sibs and maternal half-sibs.

Can you provide an explanation for why the two covariances are not equal?

2b. Assume Felipe Ruiz has calculated the covariance between daughters and dams to be 2500 and also has calculated the covariance between full sibs (same father and mother) to be 3000. Assume these are population covariances.

Can you provide an explanation for why the two covariances are not equal?

(25 points)
3. Given:

$h^2 = .40$ $\quad \sigma^2_{01} = 100$
$\sigma^2_{20} = 200$ $\quad \sigma^2_{X} = 1000$
$\sigma^2_{30} = 50$

Relatives (x and y)	a_{xy}	d_{xy}	c_{xy}
A, B	1/2	1/4	.10
C, D	1/2	0	.05
C, E	1/4	0	.20
G, H	1 + 1/4	0	0

What will be the covariance:

a) between a record of A and a record of B?

b) between a record of C and a record of D?

c) between a record of C and a record of E?

d) between a record of G and a record of H?

(2 points; included in 25)
e) What is the correlation between two records on A?

f) Can you calculate the covariance between a racord of D and a record of E from the information given? If so, do so.

4. Assume $X_S = 20$ is the average of single records of 5 full sibs (littermates) of animal V that has record $X_V = 10$.
Assume $\sigma^2_X = 400$, $h^2 = .40$, and environmental covariance between records of full sibs is 120.

a) Predict the additive genetic value of animal V.

b) What is σ^2_T?

c) What is r_{TI} for $I = b_S X_S + b_V X_V$, the index for part a)?

d) Predict the additive genetic value of V from X_V.

e) What is the r_{TI} for $I = bX_S$ for $b = .10$?

Second Exam November, 1987

(70 points). If you are not sure of how to solve any part numerically, please show symbolically (and numerically as much as possible) how to work the part.

1. For three traits in order of trait 1, 2, and 3:
the variances (diagonals) and covariances (offdiagonals) are:

phenotypic			genetic		
100	10	30	50	-10	30
10	400	60	-10	100	60
30	60	900	30	60	200

If you are confused by this description of the variances and covariances, check with a TA.

If $v_1 = 2$ and $v_2 = 3$, the selection index using only X_1 and X_2 is $I = .6316X_1 + .6842X_2$.

a) What are the relative economic values for traits 1, 2, and 3?

b) What is the variance of the index?

c) What is the variance of T?

d) What is the expected superiority in $T = v_1G_1 + v_2G_2$ if the best 1 of 10 is selected based on I?

e) What is the expected correlated change in G_3 if selection is based on I ?

1. (continued)

f) Given for an animal that: $X_1 = 10$, $X_2 = 20$, and $X_3 = 30$; what is the probability that T is greater than zero using I ?

A stubborn friend of mine thinks that an easier method of selection is to use $I_3 = 4X_3$.

g) What is the variance of I_3?

h) What is the expected superiority in $T = v_1G_1 + v_2G_2$ if the best 1 of 10 is selected based on I_3?

i) What is the expected correlated change in G_3 if selection is based on I_3?

Assume if $I = .6316X_1 + .6842X_2$ is used that the generation interval is 4 years and if $I_3 = 4X_3$ is used that the generation interval is 2 years for both males and females.

j) Calculate expected gain per year in $T = 2G_1 + 3G_2$ for the two indexes with selection of the best 1 of 10.

(30 points) Problem 2 is a single trait problem.

2. Assume:

Phenotypic variance is	100
Heritability is	.25
Dominance variance is	10
Environmental covariance between full sibs is	5
Environmental covariance between animals that mate is	20
Environmental covariance between parent and progeny is	0

a) Show clearly how to find the selection index to predict the <u>average additive genetic value</u> of two full sibs. One of the full sibs has a record (X_1). The average of single records of the two parents (X_2) is also available. Do not solve for the weights!!!

2. (continued)

b) Show clearly how to find the selection index to predict the <u>additive genetic value</u> for the full sib that has a record using X_1 and X_2. Do not solve for the weights!!!.

c) Indicate clearly how you would change either part a) or part b) to predict <u>the record</u> of the full sib that <u>does not</u> have a record. Do not solve for the weights!!!

Final Exam December, 1987

1. (25 points) This is a single trait problem.

Animal D is the dam of animal C.

The average of 4 records of animal D is $X_2 = +12$.

Animal Z is the sire of animal C.

Animals A and B are full sibs.

The average of single records of animals A and B is $X_1 = +8$.

The parents of A and B are animals S and D.

The inbreeding coefficient of animal C is $F_C = .25$.

Assume:

$\sigma^2_{10} = 20$ $\sigma^2_{01} = 16$ $\sigma^2_{20} = 8$ $\sigma^2_{PE} = 16$ $\sigma^2_X = 100$

The environmental covariance between full sibs is .04(100) = 4.

a) Show clearly how to predict the additive genetic value of animal C. Do not solve any equations.

b) Show clearly how to predict a future record of C.

c) What is σ^2_T for part a) and for part b)?

d) Show clearly how to compute r^2_{TI} for part a) and for part b).

e) What is heritability for this trait?

f) What is repeatability for this trait?

g) What is the expected covariance between a record of C and a record of a progeny of C?

h) What is the expected correlation between a record of animal C and another record of animal C?

(25 points) This is a two trait problem.

2. Records are available for trait one (growth rate) on an animal (X_1) and on trait two (loin eye area after slaughter) on a full sib (X_2).

The phenotypic variances are: for trait one, 4.00; and for trait two, .50. The phenotypic correlation is .80 between trait one and trait two on the same animal.

The heritabilities are: for trait one, .40; and for trait two, .60. The genetic correlation between the genetic value for trait one and trait two on the same animal is .60.

Assume the economic values are: for trait one, 6; and for trait two, 12.

a) Show clearly how to find the selection index to predict additive genetic value for trait one. Use trait one measured on the animal (X_1) and trait two measured on its full sib (X_2). Do not solve for the weights!!!

b) Show clearly how to find the selection index to predict additive genetic value for trait two. Use trait one measured on the animal (X_1) and trait two measured on its full sib (X_2). Do not solve for the weights!!!

2. (Continued)

c) Show clearly how to find the selection index to predict $T = 6$ (additive genetic value for trait one) + 12 (additive genetic value for trait two). Use trait one measured on the animal (X_1) and trait two measured on its full sib (X_2). Do not solve for the weights!!!

d) Show clearly what needs to be changed in c) if X_1 is trait one and X_2 is the record for trait two both measured on the animal being evaluated, with $T = 6G_1 + 12G_2$, where G_i is the additive genetic value for trait i. The full sib has no records. Do not solve for the weights!!!

(25 points) This is a one trait problem.

If you cannot work a part of this problem that is required for a later part, indicate clearly symbolically (and numerically as much as possible) how to do the later part.

3. Animal G has an adjusted record of $X_G = 20$. A maternal half-sib, H, has a record of $X_H = -16$. (G and H have the same mother and different fathers.)

Assume $\sigma_X^2 = 400$, $h^2 = .25$, $c_{GH} = .1875$

a) Predict the additive genetic value of G using X_G and X_H. You will need to solve for the selection index weights.

b) What is the r_{TI} for this index (a)?

3. (Continued)

c) What is the probability that the additive genetic value of G is greater than 8?

d) An approximate index is to weight the records by the relationships between the animal with the records and the animal being evaluated, that is,

$$I^* = a_{1\alpha}h^2X_1 + a_{2\alpha}h^2X_2 \quad .$$

For an animal and its maternal half-sib, the approximate index is:

$$I^* = .25(20) + (.25)(.25)(-16) = 4$$

What is the r_{TI} for this index (I^*)?

e) What is the correlation between the index from part a) and the index from part d)?

(25 points)

4. Cytoplasmic effects are thought to be passed directly from mother to daughter. Thus, animals that trace directly from female to female to the same female will have cytoplasmic effects alike. Assume the variance of cytoplasmic effects is σ_c^2.

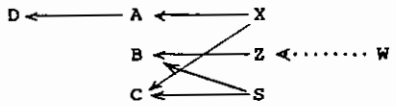
a) For the following set of data, set up the mixed model equations to predict the additive genetic value for each animal with a record and also the cytoplasmic effect for each animal, A, B, C, and D. Pretend all animals are unrelated. All animals with records are females. Assume management is a fixed effect on the records.

Animal	Management	Source of cytoplasm	Record
A	Galton	X	120
B	Galton	W	110
C	Pollak	X	90
D	Pollak	A	80

Assume $h^2 = .40$, $r = .60$, $\sigma_c^2 = .10\sigma_y^2$, $\sigma_y^2 = 400$.

4. (Continued)

b) Now assume the diagram describes the relationships among the animals.



Show clearly (symbolically and numerically as much as possible) how to modify part a) to set up the mixed model equations to predict additive genetic values for animals A, B, C, D, S, and X and the cytoplasmic values for A, B, C, D, X, and W.