THE GENETIC BASIS FOR CONSTRUCTING SELECTION INDEXES[1]

L. N. HAZEL[2]
Iowa State College, Ames, Iowa

Received May 27, 1943

"The key is man's power of accumulative selection: nature gives successive variations; man adds them up in certain directions useful to him."—Darwin, p. 35, sixth edition of *The Origin of Species*. 1920.

INTRODUCTION

THE idea of a yardstick or selection index for measuring the net merit of breeding animals is probably almost as old as the art of animal breeding itself. In practice several or many traits influence an animal's practical value, although they do so in varying degrees. The information regarding different traits may vary widely, some coming from an animal's relatives and some from the animal's own performance for traits which are expressed once or repeatedly during its lifetime. LUSH (1935) emphasized that permanent improvement from phenotypic selection is proportional to the additively genetic (heritable) fraction of the observed variance and that this varies for different traits. DOBZHANSKY (1937) suggested "that most, and possibly all, genes have manifold effects." These factors make wise selection a complicated and uncertain procedure; in addition fluctuating, vague, and sometimes erroneous ideals often cause the improvement resulting from selection to be much less than could be achieved if these obstacles were overcome.

In the initial stages of breeding investigations conducted by the REGIONAL SWINE BREEDING LABORATORY and cooperating state experiment stations, an arbitrary method of selecting breeding animals had to be adopted. In the meantime the theoretical aspects of the problem were investigated while data were being collected. While many fundamental genetic problems are still incompletely solved (particularly as regards the prevalence of dominance, epistasis and pleiotropic effects of genes in quantitative inheritance, the nature of heterosis, and the interaction of genotype and environment), the accumulation of genetic knowledge justifies an exploration of this problem. According to formulas presented by HAZEL and LUSH (1943), selection for an index which gives proper weight to each trait is more efficient than selection for one trait at a time or for several traits with an independent culling level for each trait. The principles of constructing and using selection indexes which permit the attainment of maximum genetic progress are given in the present paper. Three selection indexes were constructed (and compared as to efficiency) from data taken on the Iowa Station swine herd from the fall of 1937 through the spring of 1940. Such indexes are subject to revision when the accuracy of

---

[1] Journal paper No. J-1121 of the IOWA AGRICULTURAL EXPERIMENT STATION, Ames, Iowa. Project No. 32.

This research was done in cooperation with the REGIONAL SWINE BREEDING LABORATORY, BUREAU OF ANIMAL INDUSTRY, U. S. DEPARTMENT OF AGRICULTURE.

[2] Now at the WESTERN SHEEP BREEDING LABORATORY, Dubois, Idaho.

GENETICS 28: 476 November 1943

the statistics upon which they are based can be increased by analyzing additional data.

Selection indexes, constructed with attention to the genetic and economic bases for the various traits, should be valuable in livestock breeding programs. GALTON'S "Law of Regression," presented before the Mendelian nature of inheritance was clear, represents an early step in index construction. The sire index, widely used in selecting sires for butterfat and milk production, is a practical example of an index based on one trait but using information about several relatives. SMITH (1937) developed an index designed for the selection of plant lines, using FISHER'S concept of discriminant functions to derive a linear equation based on observable characteristics as the best available guide to the genetic value of each line.

ANALYTICAL METHOD

The net genetic improvement which can be brought about by selecting among a group of animals is the sum of the genetic gains made for the several traits which have economic importance. It is logical to weight the gain made for each trait ($\overline{G}_i$) by the relative economic value of that trait (a_i). Thus the average genetic superiority of a selected group over the group from which it was chosen is

$$\overline{H} = a_1\overline{G}_1 + a_2\overline{G}_2 + \cdots + a_n\overline{G}_n. \tag{1}$$

The relative economic value for each trait depends upon the amount by which profit may be expected to increase for each unit of improvement in that trait. Good approximations to relative economic values often can be obtained from long-time price averages and cost-of-production figures. As an example, WINTERS (1940) found that the average price of wool per pound was 3.4 times that of lamb per pound. If additional feed or labor costs are associated with increased production for either trait, the increased cost per unit should be discounted when calculating relative economic values. These values may vary from breed to breed or from region to region within the same breed. They may change, even while a breeding program is in progress, if permanent shifts in market demand occur.

Animals vary in breeding value, as in phenotype, for each of the several traits. The aggregate value of an animal is the sum of its several genotypes (assuming a distinct genotype for each economic trait), each genotype being weighted according to the relative economic value of that trait. An animal's genotype for a given trait may be defined as the sum of the average (strictly additive) effects of its genes which influence that trait. Thus the aggregate genotype of an animal is

$$H = a_1G_1 + a_2G_2 + \cdots + a_nG_n. \tag{2}$$

Environmental factors, dominance, and epistasis may make phenotypic performance unlike the genotype for that trait; hence animals having the highest values for H cannot be recognized directly with perfect accuracy. Selection for improved breeding value therefore must be practiced indirectly by selecting directly for a correlated variable (I) based on the phenotypic performance of each animal for the several traits. When selection is practiced on a large population, the genetic average of the selected group minus that of the original population represents the genetic gain from selection ($\bar{H}$ in equation 1), being

$$\bar{H} = (\bar{\imath})R_{IH}\sigma_H. \tag{3}$$

Here $\bar{\imath}$ is the average superiority of the indexes (in standard deviation units) for the selected group as compared to the whole group, R_{IH} is the correlation between H and I, while σ_H is the standard deviation of H in the whole group.

The selection differential ($\bar{\imath}$) depends upon the proportion which can be culled, being limited by the percentage needed for replacements for a particular species. The standard deviation of breeding value,

$$\sigma_H = \sqrt{a_1^2\sigma_{G_1}^2 + a_2^2\sigma_{G_2}^2 + \cdots + 2a_1a_2\sigma_{G_1}\sigma_{G_2}r_{G_1G_2} + \cdots},$$

depends upon gene frequency and to some extent upon the mating system. These can be changed only a little by the breeder. Consequently the opportunity for increasing the progress expected from selection lies in making R_{IH} as large as possible. Accordingly I is defined as

$$I = b_1X_1 + b_2X_2 + \cdots + b_nX_n, \tag{4}$$

where the X's represent phenotypic performance for the several traits and the b's are multiple regression coefficients chosen so as to make R_{IH} as large as possible. These regression coefficients may be calculated from n simultaneous equations

$$\begin{array}{llll}
\beta_1 & + \beta_2 r_{X_1X_2} & + \cdots + \beta_n r_{X_1X_n} & = r_{X_1H} \\
\beta_1 r_{X_1X_2} & + \beta_2 & + \cdots + \beta_n r_{X_2X_n} & = r_{X_2H} \\
\cdots & \cdots & \cdots & \cdots \\
\beta_1 r_{X_1X_n} & + \beta_2 r_{X_2X_n} & + \cdots + \beta_n & = r_{X_nH}
\end{array} \tag{5}$$

where $\beta_i = b_i \dfrac{\sigma_{X_i}}{\sigma_H}$ and r_{X_iH} is the correlation between H and the i-th phenotypic measurement.

The simultaneous equations can be solved only if estimates of the various correlations can be calculated. The usual methods of interclass and intraclass correlation are generally sufficient to calculate the phenotypic correlations ($r_{X_iX_j}$). WRIGHT's (1934) method of path coefficients is convenient for calculating the more complex correlations between H and phenotypic performance (r_{X_iH}). The path coefficient diagram (fig. 1) indicates the various relations between H and the phenotypic measurement for each trait (X_i). This correlation is the sum of the various paths from X_i to H, as follows,

$$r_{X_iH} = r_{G_iX_i}\{d_1r_{G_1G_i} + d_2r_{G_2G_i} + \cdots + d_nr_{G_nG_i}\}, \tag{6}$$

where $$d_i = a_i \frac{\sigma_{G_i}}{\sigma_H} \cdot$$

Therefore we must have estimates of genetic variability for each trait (σ_{G_i}), the correlation between genotype and phenotypic performance for each trait ($r_{X_iG_i}$), and the correlation between genotypes for different traits ($r_{G_iG_j}$) to solve the simultaneous equations.

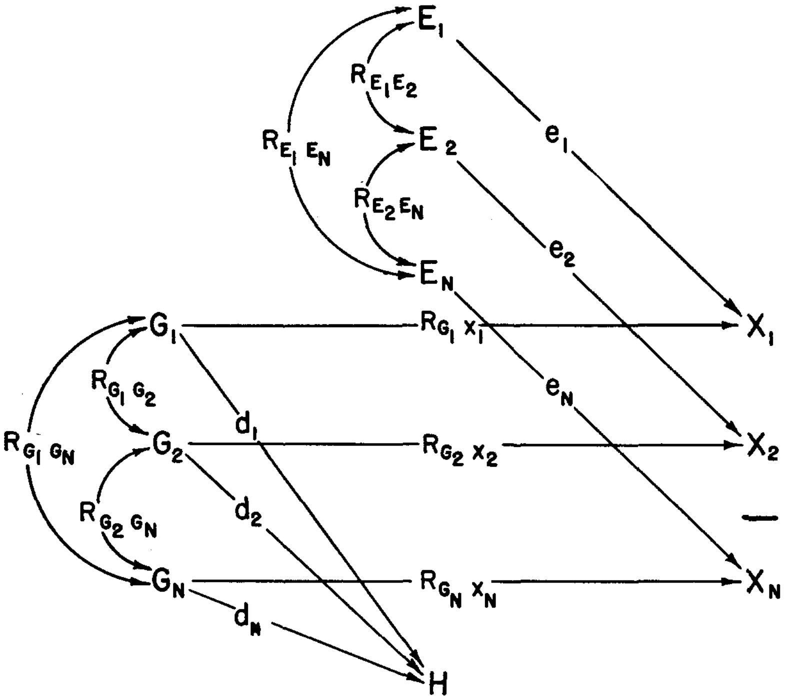

FIGURE 1. Path coefficient diagram showing the relation between phenotypic measurements (X_i) and the aggregate genotype (H). For further explanation of symbols see text.

Let an observed trait (X) be the sum of the average effects of the genes (G)[3] which an animal inherits, plus the combined effects of environment, dominance and epistasis (E); thus

(7) $$X = G + E.$$

Then the observed variance is

(8) $$\sigma_X^2 = \sigma_G^2 + \sigma_E^2,$$

---

[3] We have defined the genotype as the sum of the average effects of genes because WRIGHT (1921, 1935) has shown that only the average (additive) effects contribute to permanent gain from selection. The effects of environment, dominance, and epistasis may logically be grouped together since they act similarly to mask the genotype.

if G and E are uncorrelated, as would be the case unless particular efforts were made to give the better genotypes better or worse than average treatment. The correlation between the genotype and phenotype for the same trait is

$$r_{G_iX_i} = \frac{\sigma_{G_i}}{\sigma_{X_i}} = g_i, \tag{9}$$

where that trait is measured on the animal itself. The term g_i is the square root of the heritable (additively genetic) fraction of the observed variance.

Darwin observed the importance of correlated variation, as evidenced by the statement, "Hence if man goes on selecting, and thus augmenting, any peculiarity, he will almost certainly modify unintentionally other parts of the structure, owing to the mysterious laws of correlation." The genetic basis for this statement is evident if genes have manifold effects. Many cases of pleiotropic genes have been reported for laboratory animals, although they have received little attention in domestic animals. Linkage and non-random mating systems may also cause correlated variation; however, their effects would be less permanent and consequently less important in selection. Repeated crossing over ultimately makes the coupling and repulsion heterozygotes equally numerous, while the transient nature of breeding herds prevents sustained departure from random mating in most cases. An additional and usually much more important cause of correlated variation within an interbreeding population lies in the environmental circumstances peculiar to each animal, particularly for traits which develop during the same, or in adjacent, periods of time. Thus if trait I is correlated with trait J in the same animal, I may serve as an indicator either of the animal's genotype for J or of the environmental circumstances to which that animal was exposed when traits I and J were both being developed.

Statistically an observed phenotypic correlation may be analyzed into its constituents, a genetic correlation ($r_{G_iG_j}$) and an environmental correlation ($r_{E_iE_j}$), as indicated in figure 1. The observed phenotypic correlation between two traits measured on the same animal is

$$r_{X_iX_j} = g_ig_jr_{G_iG_j} + e_ie_jr_{E_iE_j}, \quad \text{where} \quad e_i = \frac{\sigma_{E_i}}{\sigma_{X_i}}. \tag{10}$$

Sex-limited traits and traits such as carcass merit cannot be measured directly on all breeding animals. Sometimes selection must be practiced before each animal's performance for every trait is known. For these reasons there is a possibility of increasing genetic progress in a breeding program by using information about the performance of relatives. Examples of how r_{X_iH} may be calculated are given in the subsequent section where X_i is (1) the individual's own performance, (2) the performance of the individual's dam, and (3) the average perfromance of a group of relatives of which the individual is a member.

APPLICATION

Three selection indexes for young boars and gilts were constructed from data taken on the Iowa Station swine herd from the fall of 1937 through the spring of 1940. The history of the herd was given by BYWATERS (1937) and more recently by WHATLEY (1942). LUSH (1940) showed that the economic return from the swine enterprise depended largely upon three traits—growth rate, market suitability, and size of litter—the relative economic values of which were $\frac{1}{3}$, 1, and 2, respectively, in the units used in this study. Growth rate was measured by weight at 180 days of age. Market suitability was measured by a numerical score given by each of several judges. The details of these measurements were given by WHATLEY for 180-day weight and by STONAKER and LUSH (1942) for score. The measure of litter size was that suggested by LUSH and MOLLN (1942),

$$P = n_0 + n_{21} + n_{56} + W_{21}/10 + W_{56}/30,$$

the n's and W's referring to the number of pigs and weight of the litter, respectively, at the ages designated by the subscripts. Young boars and gilts are not old enough to have produced litters themselves at the time first selections must be made; consequently the estimate of productivity for each young animal is based upon the litter in which it was born.

The aggregate breeding value of an animal is therefore defined as

$$H = \tfrac{1}{3}G_W + G_S + 2G_P,$$

G_W, G_S and G_P referring to the genotypes for 180-day weight, market score and productivity, respectively. The following statistics are needed to construct one or more of the three indexes:

A. Phenotypic constants
  1. Standard deviation for each of the three traits
  2. Phenotypic correlation between each pair of traits
  3. Phenotypic correlations between the traits of relatives

B. Genetic constants
  1. Heritable fraction of the variance in each trait
  2. Genetic correlation between each pair of traits.

Differences due to season and line of breeding were eliminated in calculating these statistics, as selection is usually practiced between animals born in the same season and within the same line or other interbreeding population. The phenotypic constants shown in table 1 were calculated by the usual methods of variance and covariance analysis (SNEDECOR 1940).

Several procedures were outlined by LUSH (1941) for estimating heritability (g_i^2) in farm livestock. Intrasire regressions of offspring on dam were considered preferable to correlation coefficients in the present case because of the rather intense selection which had been practiced in the Iowa Station swine

TABLE 1

Phenotypic constants necessary to construct one or more of the three indexes.

TRAIT	STANDARD DEVIATION	DEGREES OF FREEDOM
180-day weight (in pounds)	31.86	1513
Market score (in points)	4.78	1513
Productivity (in points)	9.72	282

CORRELATION BETWEEN TRAITS	CORRELATION COEFFICIENT	STANDARD ERROR
Pig's own weight and score	0.614	0.016
Pig's own weight and productivity of its dam	−.024	.059
Pig's own score and productivity of its dam	−.081	.059
Weight of litter mates	.314	.027
Score of litter mates	.204	.026
Weight and score of litter mates	.186	.026

herd for the traits under consideration. The intensity of selection was evidenced by the higher averages and lower variability for the dams as compared to their offspring. The intrasire regressions of offspring on dam are shown in table 2.

TABLE 2

Regression of each trait for the offspring on each of the three traits for the dam.

OFFSPRING'S TRAIT	DAM'S TRAIT	SYMBOL*	REGRESSION COEFFICIENT	DEGREES OF FREEDOM
Weight	Weight	bw_2w_1	0.149	232
Score	Score	bs_2s_1	.049	189
Productivity	Productivity	bp_2p_1	.081	82
Weight	Score	bw_2s_1	.153	189
Score	Weight	bs_2w_1	.013	232
Weight	Productivity	bw_2p_1	−.067	204
Productivity	Weight	bp_2w_1	.004	52
Score	Productivity	bs_2p_1	−.024	204
Productivity	Score	bp_2s_1	.041	52

* The sub-subscripts 1 and 2 refer to dam and offspring, respectively, the regressions being those of offspring's trait on dam's trait in each case.

Values of g_i, calculated from the formula,[4] $g_i=\sqrt{2b_{i_2i_1}}$, are given in table 3 for the three traits.

---

[4] This formula was derived from the relation,

$$b_{i_2i_1}=g_i^2t+e_i^2r_{EE}.$$

The genetic relationship between dam and offspring (t) was taken as 0.5, and the environmental correlation (r_{EE}) was assumed zero, since no particular efforts had been made to give parent and offspring similar or dissimilar treatment.

TABLE 3

Genetic constants derived from intrasire regressions of offspring on dam.

STATISTIC	SYMBOL	ESTIMATE
$\sqrt{\text{heritability of weight}}$	g_W	0.546
$\sqrt{\text{heritability of score}}$	g_S	.313
$\sqrt{\text{heritability of productivity}}$	g_P	.402
Genetic correlation between weight and score	$r_{G_WG_S}$	.519
Genetic correlation between weight and productivity	$r_{G_WG_P}$	.0±
Genetic correlation between score and productivity	$r_{G_SG_P}$	.0±

Procedures for estimating the genetic correlations ($r_{G_iG_j}$) have not been developed previously. Equation 10 is not suitable for this purpose because no direct method is available for separating the genetic and environmental correlations for two traits measured upon the same animal. For example, any environmental accident such as differential exposure to parasites or infection which affected growth and plumpness would either raise or lower the growth rate and score of the same animal. Thus the actually observed correlation between the two traits on the same animal conceivably could be due wholly to such environmental circumstances, or to genes which affect both traits, or to a mixture of the two causes in any proportion. To measure the genetic correlations by themselves it was necessary to correlate one trait in one animal with the other in a relative. The formula,[5]

$$r_{G_iG_j} = \sqrt{\frac{b_{i_2j_1} \cdot b_{j_2i_1}}{b_{i_2i_1} \cdot b_{j_2j_1}}} = \sqrt{\frac{(\text{cov } I_2J_1)(\text{cov } J_2I_1)}{(\text{cov } I_2I_1)(\text{cov } J_2J_1)}},$$

was adopted because it appeared to be unbiased by selection and to utilize most of the available information. Estimates of the genetic correlations between each pair of the three traits are given in table 3. Consistent estimates could not be derived for $r_{G_WG_P}$ and $r_{G_SG_P}$, since in each case the two regressions required in the numerator differed in sign. Since these regressions are small and not significantly different from zero, it seemed more accurate to assign both $r_{G_WG_P}$ and $r_{G_SG_P}$ values of zero than to attempt an alternative method of calculating them.

The genetic constants probably include a small fraction of the epistatic deviations and exclude the average differences between groups of dams to

---

[5] Selection for the independent variable does not bias the corresponding regression coefficient (EISENHART 1939). The quantity $\sqrt{b_{i_2j_1} \cdot b_{j_2i_1}}$ consequently can be taken as an estimate of the correlation between two traits of dam and offspring in an unselected population. The biometric relation,

$$r_{i_2j_1} = r_{j_2i_1} = g_i g_j t r_{G_iG_j},$$

permits the derivation of the formula given above for $r_{G_iG_j}$. The two regressions (or correlations) actually provide two independent estimates of $r_{G_iG_j}$ in unselected populations and an arithmetic average of the two estimates would be less biased by sampling errors than the geometric average used here.

which sires were mated. Since some inbreeding had occurred, they may be slightly nearer zero than would be expected in a non-inbred population. If they are biased, and this bias is equal and in the same direction for the different genetic constants, the index will be no less accurate for selecting breeding animals.

Weight at 180 days of age and market score are available for each animal before breeding age. The first index constructed was based on these two traits. Values for r_{WH} and r_{SH} were calculated by substituting the statistics from tables 1 and 3 into formula 6 as follows:

$$r_{WH} = \frac{0.546}{\sigma_H}\{5.803(1) + (1.496)(.519)\} = 3.593/\sigma_H$$

$$r_{SH} = \frac{0.313}{\sigma_H}\{5.803(.519) + 1.496(1)\} = 1.409/\sigma_H.$$

The two simultaneous equations are:

$$\beta_W + 0.614\beta_S = 3.593/\sigma_H$$
$$0.614\beta_W + \beta_S = 1.409/\sigma_H.$$

Solving, we get $\beta_W = 4.381/\sigma_H$ and $\beta_S = -1.282/\sigma_H$.
Since $\beta_i = b_i\sigma_{Xi}/\sigma_H$, the first index is

$$I_1 = 0.137W - 0.268S,$$

where W and S represent the pig's own 180-day weight and market score, respectively. If the index can thus be made more convenient for use or for keeping records, it may be multiplied by any constant or any constant may be added to it without influencing its relative accuracy, since such procedure does not change R_{IH}.

A second index was constructed by using the productivity of the dam as a measure of each pig's productivity in the index, the lapse of one generation being compensated by multiplying r_{PH} by one-half as follows:

$$r_{PH} = \frac{0.402}{2\sigma_H}(7.802) = 1.566/\sigma_H.$$

The three simultaneous equations were solved as before, the second index being

$$I_2 = 0.136W - 0.232S + 0.164P.$$

The third index was designed to include information about the average weight and score of the litter in which each pig was born, in addition to the three traits in the second index. The procedure used was to consider the average weight ($\overline{W}$) and score ($\overline{S}$) of the litter as a fourth and fifth variable, expressing the necessary correlations in terms of the correlations given in table 1 and as a function of the number of pigs per litter (k). This latter step was necessary because the number upon which the average is based influences the variability of the average and the amount of information in the average concerning each pig's breeding value. The additional phenotypic correlations in terms of k and the correlations in table 1 are

$$r_{W\bar{W}} = \sqrt{\frac{1 + (k-1).314}{k}},$$

$$r_{S\bar{S}} = \sqrt{\frac{1 + (k-1).204}{k}},$$

$$r_{W\bar{S}} = \frac{.614 + (k-1).186}{\sqrt{k[1 + (k-1).204]}},$$

$$r_{S\bar{W}} = \frac{.614 + (k-1).186}{\sqrt{k[1 + (k-1).314]}},$$

$$r_{P\bar{W}} = -\ .024\sqrt{\frac{k}{1 + (k-1).314}}$$

$$r_{P\bar{S}} = -\ .081\sqrt{\frac{k}{1 + (k-1).204}}, \quad \text{and}$$

$$r_{\bar{W}\bar{S}} = \frac{.614 + (k-1).186}{\sqrt{[1 + (k-1).314][1 + (k-1).204]}}.$$

The additional correlations between H and the litter averages are

$$r_{\bar{W}H} = \frac{3.593[1 + (k-1).5]}{\sigma_H\sqrt{k[1 + (k-1).314]}} \quad \text{and}$$

$$r_{\bar{S}H} = \frac{1.409[1 + (k-1).5]}{\sigma_H\sqrt{k[1 + (k-1).204]}}.$$

The five simultaneous equations were solved as before, giving the following partial regression coefficients:

$$b_W = .098$$

$$b_S = -\ .165$$

$$b_P = \frac{0.358\{0.164(k-1) + 0.016(k-1)^2\}}{2.18\{0.998(k-1) + 0.899(k-1)^2\}}$$

$$b_{\bar{W}} = \frac{k\{0.270 + (0.004)(k-1)\}}{7.14\{3.273(k-1) + 0.324(k-1)^2\}}$$

$$b_{\bar{S}} = \frac{-\ k\{0.070 + (0.02)(k-1)\}}{1.07\{0.491(k-1) + 0.049(k-1)^2\}}.$$

The absolute values of the three latter coefficients are given in table 4 for different values of k. The number of pigs in the litter has so little effect on

TABLE 4

Partial regression coefficients for dam's productivity, average weight of the litter and average score of the litter for different numbers of pigs in the litter, in the third index.

NUMBER OF PIGS IN THE LITTER (k)	b_P	$b_{\bar{W}}$	$b_{\bar{S}}$
1	0.164	0.038	−0.067
2	.165	.059	− .113
3	.165	.072	− .148
4	.166	.081	− .175
5	.166	.088	− .197
6	.167	.093	− .215
7	.167	.098	− .230
8	.167	.101	− .243
9	.168	.104	− .254
10	.168	.106	− .264
11	.168	.109	− .273
12	.169	.110	− .280
13	.169	.112	− .287
14	.169	.113	− .293
15	.169	.115	− .299

b_P that this could be taken as a constant (0.166 or 0.167) with very little error.

The amount of genetic progress expected when a given index is used in making selections is proportional to R_{IH} (see formula 3). Hence these values provide a basis for choosing an index which is easy and simple to use yet which is of nearly maximum accuracy. For example the three indexes previously constructed may be compared as follows:

$$R_{I_1H} = \sqrt{\beta_W r_{WH} + \beta_S r_{SH}} = 0.363$$

$$R_{I_2H} = 0.395 \quad \text{and}$$

$$R_{I_3H} = 0.404, \quad \text{for} \quad k = 5.$$

The second and third indexes are 8.8 and 11.3 percent, respectively, more efficient than the first. Since the time and effort expended in keeping records is but a small fraction of the total labor connected with a breeding program, the second index would almost certainly be preferable to the first. The third might also be chosen over the second, since genetic progress could be increased still further, and the extra labor would be only that of computing and using the litter averages from data already taken.

DISCUSSION

FISHER'S (1930) "fundamental theorem of natural selection" and WRIGHT'S (1931) emphasis that genetic change depends upon genetic variability and selection intensity indicate that the factors which are important in natural selection also hold for selection as practiced by man. In equation 3, σ_H is a measure of genetic variability and $\bar{\imath}$ a measure of selection intensity. The breeder in applying artificial selection to an animal population has the opportunity of increasing the accuracy of his selections (increasing R_{IH}) considerably over what may hold in the "trial-and-error" methods of natural selection. The breeder has additional opportunities of increasing $\bar{\imath}$ (within limits) and of increasing genetic variability by the intentional control of population size (inbreeding) and migration (outcrossing) which have not been considered in the present study. WRIGHT (1940) has examined the conditions under which the supplementary use of these latter methods may be advantageous.

For the special case where the traits are uncorrelated, R_{IH} is a maximum when each regression coefficient is equal to (or proportional to) the product of the relative economic value and heritability for each trait ($b_i = a_i g_i^2$). The correlations which may exist between traits complicate the calculation of the partial regression coefficients, just as correlations between the independent variables do in any multiple regression analysis. When the phenotypic correlation is large as compared to the genetic correlation, the regression coefficient for a trait with little economic importance or slight heritable variation may be negative, because its function in the index then becomes mainly that of indicating the environment for a more important and more highly heritable trait. An example of this was seen in the negative regression coefficients for score in the three indexes calculated previously. However, selection for the animals having the highest indexes would create some improvement for score because of the positive genetic correlation between score and weight.

An index constructed from data taken on a herd in one locality may not be widely applicable. The reasons for this are:

1. Relative economic values for a trait may vary with the particular locality or nature of the enterprise.

2. The genetic constitution of herds may differ, especially where they are under distinctly non-random mating systems such as intense inbreeding.

3. Different managerial practices may cause the standard deviations for the traits to vary in different herds. The standard deviations for subjective traits such as market conformation measured by judging or by scores may vary because different judges will vary the range over which they spread their scores.

4. Few herds are large enough to provide data sufficient to make the sampling errors of the genetic constants small.

The best way to test whether or not selection indexes can be standardized and recommended for general use seems to be to compare several indexes constructed from data taken on different herds.

The data in the present study were sufficiently numerous to provide accurate estimates of the phenotypic constants in table 1. They were less satisfactory with regard to the reliability of the genetic constants given in table 3. Some idea of the general accuracy of these figures can be obtained by comparing them to similar estimates by other investigators. WHATLEY (1942) used several methods to estimate heritability of 180-day weight in the Iowa Station swine herd through 1938, concluding that "at least 30 percent and possibly more than 40 percent of the individual variance" was hereditary. WHATLEY and NELSON (1942) estimated that 180-day weight in the Oklahoma Duroc herd was 23 percent heritable, while BAKER and coworkers (1943) found a figure of 25 percent for 168-day weight in the Nebraska (North Platte substation) Duroc herd. These estimates help to substantiate the figure of 30 percent for the heritability of 180-day weight found in this study. WHATLEY and NELSON also estimated that market score was about 33 percent heritable, while STONAKER and LUSH (1942) obtained an estimate of 20 percent for the Iowa Station herd from data which included that used in this study. These estimates indicate that our estimate of 10 percent for the heritability of score may be too small. Previous reports of the heritability of productivity have not been made, but LUSH and MOLLN (1942) found that the correlation between litters by the same sow was between 0.15 and 0.20 for the items which are included in productivity. They quoted a number of investigations which substantiated their findings. In general these results substantiate our figure of 16 percent if most of the permanent differences in these items between sows are hereditary.

Other attempts to estimate genetic correlations have not been made. The genetic correlation of 0.52 between 180-day weight and score indicates that about half of the genes which influence one trait also influence the other (assuming equal gene frequency and equal effects of the genes). However most of the observed correlation of 0.61 was due to an environmental correlation. Although an effort was made to score all pigs at a constant weight of 225 pounds, part of this correlation was probably due to a subjective tendency on the part of the judges to assign scores in accordance with differences in age or weights of the pigs. The genetic correlations between productivity and weight or score may have been either positive or negative so far as the evidence from the present data indicate. They appeared to be small; hence assigning them a value of zero is unlikely to have caused serious errors in the indexes.

From the studies of heritability which have been made for economic traits in different farm animals, it seems that the best indexes which can be constructed will be far from perfect. The confusing effects of environment, dominance, and epistasis in masking genotypes cause the progress in the present case to be less than half of what might be made if genotypes could be recognized precisely. Thus the indexes constructed for swine permit from 36.1 to 40.4 percent as much gain as could be made with a perfect index (where $R_{IH}=1$), which is the limit of what could be achieved if the exact Mendelian composition of every animal were known. These indexes could be improved somewhat by more perfect control of the environment, by the wise use of corrections for known environmental circumstances, by more accurately measuring differences in phenotypes, and by including the performance of additional relatives in the index; however, the use of these methods is limited by practical considerations. Although R_{IH} is likely to increase with the age of the animal (as more becomes known about its phenotype and as its progeny becomes observable) so that more gain can be made from selecting within a group of older animals, the length of generation will also increase. DICKERSON and HAZEL (1942) have shown that the interval between generations in some cases is increased by progeny testing more than enough to offset the increased accuracy of selection, the net result of more emphasis on the progeny test then being a decrease in the *annual* rate of genetic improvement. While these considerations do not indicate much possibility of phenomenally rapid improvement in animal populations from selection alone, the progress which can be made with properly constructed indexes is considerably greater than can be expected when the ideals toward which selection is directed are confused or erroneous.

CONCLUSIONS

The genetic gain which can be made by selecting for several traits simultaneously within a group of animals is the product of (1) the selection differential, (2) the multiple correlation between aggregate breeding value and the selection index, and (3) genetic variability. The first of these may be very small due to the breeder's carelessness, procrastination, etc., and is limited by the rate of reproduction for each species, while the third is relatively beyond man's control; hence the greatest opportunity of increasing the progress from selection is by insuring that the second is as large as possible.

A multiple correlation method of constructing selection indexes having maximum accuracy was presented. The following constants must be known in order to solve the simultaneous equations:

1. Relative economic values for the different traits
2. Phenotypic constants
   a. Standard deviations for each trait
   b. Correlation between each pair of traits
3. Genetic constants
   a. Heritability of each trait
   b. Genetic correlations between each pair of traits

Examples of the construction of selection indexes for young boars and gilts were presented from data taken on the Iowa Station swine herd using (1) 180-day weight and market score of the individual animal, (2) the two previous traits and productivity of the dam, and (3) the three previous traits and the average weight and score of the litter of which each pig is a member.

The progress which can be made by using the above indexes varied from 36 to 40 percent of that which could be made with a perfect index. The loss is due to the confusing effects of environmental circumstances, dominance, and epistasis, all of which can make phenotypes unlike genotypes.

ACKNOWLEDGMENT

A formal statement of appreciation is scarcely indicative of the author's indebtedness to DR. JAY L. LUSH for the use of many of his unpublished notes, his inspiration and guidance. The author is also indebted to PROF. W. G. COCHRAN, whose suggestions regarding statistical procedure were invaluable.

LITERATURE CITED

Baker, M. L., L. N. Hazel, and C. F. Reinmiller, 1943 The relative importance of heredity and environment in the growth of pigs at different ages. J. Anim. Sci. **2**: 1–13.

Bywaters, J. H., 1937 The hereditary and environmental portions of the variance in weaning weights of Poland-China pigs. Genetics **22**: 457–468.

Dickerson, G. E., and L. N. Hazel, 1942 Effect of progeny testing on improvement expected within closed herds. (In process of publication.) Abstract in J. Anim. Sci. **1**: 342.

Dobzhansky, Th., 1937 Genetics and the origin of species. New York: Columbia University Press.

Eisenhart, C., 1939 The interpretation of certain regression methods and their use in biological and industrial research. Ann. Math. Statist. **10**: 162.

Fisher, R. A., 1930 Genetical theory of natural selection. London: Oxford University Press.

Hazel, L. N., and J. L. Lush, 1943 The efficiency of three methods of selection. J. Hered. **33**: 393–399.

Lush, J. L., 1935 The inheritance of productivity in farm livestock. Pt. V. Discussion of preceding contributions. Emp. J. Exp. Agric. **3**: 25–30.

1940 Relative importance of dam's productivity, pig's own market score and pig's own growth rate, in selecting young boars and gilts for breeding. Research Item No. 18 of the Regional Swine Breeding Laboratory. Mimeographed. Ames, Iowa.

1941 Intra-sire correlations or regressions of offspring on dam as a method of estimating heritability of characteristics. Proc. Amer. Soc. Anim. Prod. for **1940**: 293–301.

Lush, J. L., and A. E. Molln, 1942 Litter size and weight as permanent characteristics of sows. Tech. Bull. U. S. Dept. Agric. No. **836**.

Smith, F., 1937 A discriminant function for plant selection. Ann. Eugen. **7**: 240–250.

Snedecor, G. W., 1940 Statistical methods. Third edition. Ames, Iowa: Iowa State College Press.

Stonaker, H. H., and J. L. Lush, 1942 Heritability of conformation in Poland-China swine as measured by scoring. J. Anim. Sci. **1**: 99–105.

Whatley, J. A., Jr., 1942 Influence of heredity and other factors on 180-day weight in Poland-China swine. J. Agric. Res. **65**: 249–264.

Whatley, J. A., Jr., and R. H. Nelson, 1942 Heritability of differences in 180-day weight and market score of Duroc swine. (Abstract.) J. Anim. Sci. **1**: 70.

Winters, L. M., 1940 Records of performance for meat animals. Emp. J. Exp. Agric. **8**: 259–268.

Wright, S., 1921 Systems of mating. Genetics **6**: 111–178.

1931 Evolution in Mendelian populations. Genetics **16**: 97–159.

1934 The method of path coefficients. Ann. Math. Statist. **5**: 161–215.

1935 The analysis of variance and the correlation between relatives with respect to deviations from an optimum. J. Genet. **30**: 243–256.

1940 Genetic principles governing the rate of progress of livestock breeding. Proc. Amer. Soc. Anim. Prod. for **1939**: 18–26.