Principles of a selection index which involves several characteristics and utilizes information concerning relatives

Lanoy Nelson Hazel

Iowa State College

Follow this and additional works at: https://lib.dr.iastate.edu/rtd
Part of the Agriculture Commons, Animal Sciences Commons, and the Genetics Commons

Recommended Citation
Hazel, Lanoy Nelson, "Principles of a selection index which involves several characteristics and utilizes information concerning relatives " (1941). Retrospective Theses and Dissertations. 12775.
https://lib.dr.iastate.edu/rtd/12775

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
PRINCIPLES OF A SELECTION INDEX WHICH INVOLVES
SEVERAL CHARACTERISTICS AND UTILIZES
INFORMATION CONCERNING RELATIVES

by

Lancy Nelson Hazel

A Thesis Submitted to the Graduate Faculty
for the Degree of
DOCTOR OF PHILOSOPHY
Major Subject Animal Breeding - Genetics

Approved:

Signature was redacted for privacy.

In charge of Major work

Signature was redacted for privacy.

In charge of Major work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State College
1941
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>4</td>
</tr>
<tr>
<td>II. HISTORICAL</td>
<td>8</td>
</tr>
<tr>
<td>A. The Method of Path Coefficients</td>
<td>8</td>
</tr>
<tr>
<td>B. The Roles of Heredity and Environment</td>
<td>9</td>
</tr>
<tr>
<td>C. Previous Indexes</td>
<td>14</td>
</tr>
<tr>
<td>III. THE INVESTIGATION</td>
<td>17</td>
</tr>
<tr>
<td>A. Procedure</td>
<td>17</td>
</tr>
<tr>
<td>B. Theoretical Investigation</td>
<td>18</td>
</tr>
<tr>
<td>1. The factors governing rate of genetic change</td>
<td>18</td>
</tr>
<tr>
<td>2. Genetic basis for correlations between genotype and phenotype</td>
<td>22</td>
</tr>
<tr>
<td>3. Principles of index construction</td>
<td>31</td>
</tr>
<tr>
<td>a. The selection differential</td>
<td>31</td>
</tr>
<tr>
<td>b. The aggregate genotypic variability</td>
<td>34</td>
</tr>
<tr>
<td>c. The correlation between the aggregate genotype and the index</td>
<td>35</td>
</tr>
<tr>
<td>4. Interpretation of theoretical deductions</td>
<td>47</td>
</tr>
<tr>
<td>a. Methods of selection</td>
<td>48</td>
</tr>
<tr>
<td>b. Genetic progress by indirect selection</td>
<td>52</td>
</tr>
<tr>
<td>C. Application of the Index to an Actual Population</td>
<td>57</td>
</tr>
<tr>
<td>1. Source and adjustment of data</td>
<td>57</td>
</tr>
<tr>
<td>2. Constants necessary for the construction of the index</td>
<td>60</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>3.</td>
<td>Method of analysis</td>
</tr>
<tr>
<td>4.</td>
<td>Results of the analysis</td>
</tr>
<tr>
<td></td>
<td>a. Phenotypic constants</td>
</tr>
<tr>
<td></td>
<td>b. Genotypic constants</td>
</tr>
<tr>
<td>5.</td>
<td>The regression coefficients</td>
</tr>
<tr>
<td>6.</td>
<td>Test of the index</td>
</tr>
<tr>
<td>7.</td>
<td>The comparison of indexes</td>
</tr>
<tr>
<td>IV.</td>
<td>DISCUSSION</td>
</tr>
<tr>
<td>V.</td>
<td>SUMMARY AND CONCLUSIONS</td>
</tr>
<tr>
<td>VI.</td>
<td>LITERATURE CITED</td>
</tr>
<tr>
<td>VII.</td>
<td>ACKNOWLEDGMENTS</td>
</tr>
</tbody>
</table>
I. INTRODUCTION

Often glimpses and occasionally even revealing insight into the effects of the forces which operate to produce the natural phenomena observed in organic populations can be obtained by the construction of mathematical images to represent physical conditions. Certain hypotheses and possibly some artificial concepts must usually be invoked in order to reduce the perplexing disorganization of the actual to an orderly system capable of being comprehended by the intellect. Mendel's brilliant hypothesis concerning the genetic constitution of the pollen and ova of hybrids is a striking example of the method. When theoretical laws have been demonstrated to explain experimental results repeatedly they come to be recognized as natural or empirical laws, at least until more recent discoveries cause their modification or replacement by additional hypotheses.

The science of genetics has created a theory of the inheritance of quantitative characteristics -- briefly, differences in the expression of a single characteristic are the result of the combined effects of a large number of genes and innumerable environmental factors. The effects of either genes or environmental factors may be exerted at any time during the development of the zygote, beginning with the formation of the egg. This theory has been substantiated repeat-
edly by experimental investigations, classic examples of which are those of Johannsen, Nilsson-Ehle and Castle.

The combination of genetic theory and statistical analysis is particularly well adapted to the study of the effects of selection upon organic populations. Attempts to apply the knowledge gained from such studies to the improvement of domestic animals are not new. The widespread use of sire indexes, progeny tests and selective registration is evidence that progressive breeders have received this knowledge in its applied form with eagerness.

Nevertheless many uncertainties still attend the formulation of a constructive breeding program. Breeders and authorities in animal breeding have generally assumed that direct selection for an easily measured character could create genetic improvement in a correlated character that was measured only with difficulty or rather late in life. For example, Copeland (1941) concluded that ideals of type and production can be readily combined, interpreting a correlation of -.18±.06 between the type of the sire and the fat production of daughters as signifying "no relation at all between the classification rating of a bull and the production of his progeny." Actually the genetic basis for such an assumption has not been investigated. Little is known of the effects of simultaneous selection for two or more characteristics upon the genetic composition of populations, although it is a general rule that at least several charac-
teristics merit improvement in practically all breeds of livestock.

Genetic improvement can be induced only by selection -- by allowing genotypes to contribute to the next generation according to their relative merit. The exact value of the genotypes for quantitative characters can rarely, if ever, be determined precisely; hence, selection of the best is implied by selection of the presumably best. Methods of combining information about different characteristics and from different sources as an aid in the selection of animals have received little attention. Accurate selection is seldom a simple process; questions and uncertainties arise with the culling of almost every animal where a choice of animals is involved. Some of the most important of the confusing features are:

(1) The relative importance of the several economic characteristics differs

(2) The bits of knowledge available about the genotypes of the several characteristics of the same animal are not equally informative

(3) The information available on different animals varies for one or more characteristics

Probably most breeders form rather general impressions regarding these points, but the choice of breeding animals must be influenced considerably by the quest for fancy points, unconscious favoritism or fluctuating ideals -- factors which
would not influence an "index" based on the objective evaluation of the pertinent information available concerning each animal in the herd.

The purposes of this study were: (1) To investigate the factors which control the rate of genetic change caused by simultaneous selection for several characteristics and (2) to develop the principles of an index, the use of which will permit the attainment of maximum genetic change by selection.
II. HISTORICAL

The statistical study of organic populations, originated by Galton and extended by Pearson, was greatly aided by the development of correlation and multiple regression techniques. Although the rediscovery of Mendelism necessitated the reinterpretation of many theories proposed as a result of the early statistical investigations, application of the statistical techniques themselves has spread to many fields of science and proved increasingly useful. The methods are so well known as to require no elaboration here.

More recently methods of combining genetic theory and statistical procedure have been developed. As a result, a great body of facts and theories in the field of population genetics has been amassed. A review of the contributions pertinent to the arguments developed in the present problem must include much of method as well as of genetic theory from the standpoint of populations.

A. The Method of Path Coefficients

Wright (1921a, 1934) proposed the method of path coefficients as an aid in the quantitative analysis of correlations connected by a system of causal effects. The applications of the method extend to cases where the causal rela-
tions of a system of variables are known or can be assumed as an hypothesis. In cases where no causal relations are assumed the method is identical with that of multiple regression. Specifically, the path coefficient is that fraction of the standard deviation of the dependent variable (including the appropriate sign) for which the designated factor is directly responsible.

The two important properties of path coefficients, (1) the analysis of correlations and (2) the measurement of determination, are illustrated by fig. 1. The second property corresponds to the analysis of variance which is more direct where tests of significance are desired. The principal difficulty in the application of the path coefficient method lies in distinguishing relations of direct cause and effect from those which are due to common causes. In this connection Wright states, "the formulation of hypotheses is emphatically the business of one who is thoroughly familiar with the realities of the case."

B. The Roles of Heredity and Environment

Wright (1920) used path coefficients in studying the roles of heredity and environment in the development of characteristics and measured the relative importance of the two in determining the piebald pattern in guinea pigs. The concept has been amplified more recently by Wright (1927, 1937)
Fig. 1. Path coefficient diagram to illustrate the analysis of an observed correlation between X and Y (I). A, B, C and D are causative factors which completely determine the values of X and Y (II). The path coefficients are represented by small letters.
and Fisher (1930). According to the more recent concept differences in the phenotypes of individuals may be due either to genetic or environmental differences. The genotype is construed as the sum of the average effects of the individual genes which influence the expression of a given character. Dominance deviations from this additive scheme arise because the effect of a gene is sometimes greater or less than average, depending upon whether or not an allelicomorph like it is already present. Epistatic deviations result whenever the size or direction of the effect of a gene varies according to which nonallelicomorphic genes are present.

According to the preceding scheme the variance observed in a characteristic can be analyzed into several parts,

\[ \sigma^2_X = \sigma^2_G + \sigma^2_D + \sigma^2_E, \]

where \( \sigma^2_X \) = total variance,
\( \sigma^2_G \) = additively genetic variance,
\( \sigma^2_D \) = variance due to dominance deviations,
\( \sigma^2_I \) = variance due to epistatic interactions,
\( \sigma^2_E \) = variance due to environmental differences.

This analysis is based upon the assumption that genetic and environmental effects are linear and independent of each other. The path coefficient analysis of these relations (fig. 2) permits the derivation of the following formulae:
Fig. 2. Path coefficient diagram to illustrate the analysis of differences between phenotypes (observed characteristics) which are completely determined by differences in

- $E = \text{environment}$
- $D = \text{dominance deviations}$
- $I = \text{epistatic interactions}$
- $G = \text{average gene effects}$
\[ 1 = g^2 + i^2 + d^2 + e^2, \]

and

\[ r_{gx} = g. \]

Strictly speaking, the dominance and epistatic deviations are genetic, but since their action parallels the action of environmental differences in making the genotype unlike the phenotype and are not transmissible (except for a small fraction of the epistatic combinations), it is sufficient for many purposes to treat them as one with the environmental differences. For example, Wright (1920) found that the additively genetic fraction of the variance in color pattern in a random-bred stock of guinea-pigs was .42, leaving .58 as the unanalyzed fraction from other sources. Independent confirmation of the theory was obtained from a highly inbred stock, the variance of which was only .57 as large as that of the random-bred stock.

Wright (1921b) showed that the mating system influenced the genetic fraction of the variance. Fisher (1930) demonstrated that the intensity of selection was the primary agency in changing gene frequency. Since either of these factors may differ widely from herd to herd, the results of an analysis from one population can be applied to other populations only after careful examination of their fundamental genetic similarity.

Lush (1935) emphasized the importance of the additively genetic fraction of the variance only as being subject to simple mass selection.
Wright (1940) investigated the effects of dominance, factor interaction and gene frequency upon the correlations between parent and offspring, between full sibs and upon the additive portion of the variance. He discussed the implications of these factors as they affected the rate of genetic progress by selection and suggested modifications of the breeding program according to the existing genetic situation.

Lush (1941) examined the various methods of estimating the additively genetic fraction of the variance and indicated the advantages and disadvantages of each which might be encountered in data taken from farm livestock.

C. Previous Indexes

The first index (as the term is used in this study) was devised by Galton (1889) in attempting to predict height in man from a knowledge of the height of relatives -- the first application of the now well-known "Law of Regression." Indexes have since been devised for several types of problems in animal breeding. Sire indexes as a means of evaluating the breeding value of dairy sires by the production of their offspring have been most widely popularized.

The principle of equalizing the effects of differences in the average merit of females to which different males happen to get mated has been recognized by many workers
in comparing the breeding values of males by means of the progeny test. It was used by Hansson for dairy bulls in 1913. Schmidt (1919) introduced "diallel crossing," which is based on this principle, into genetic literature. Yapp (1925) suggested that twice the average performance of the daughters minus the average performance of the dams was an accurate and useful measure of the transmitting ability of dairy sires. Goodale (1927) suggested a revision of the former index based on the assumption that genes which affected milk production in the plus direction were dominant. Lush (1933) criticized Goodale's suggested revision upon the basis that offspring average midway between their parents in populations mating at random.

From a study of the theoretical consequences of Mendelian inheritance, Wright (1932a) developed the following formula for estimating the breeding value of a sire:

\[ S = A + \frac{n}{n+2} (\bar{D} - \bar{D} - \bar{A}), \]

\( n \) being the number of tested daughters, \( A \) the breed average, \( \bar{D} \) the average production of the daughters and \( \bar{D} \) the average production of the dams. He pointed out the fact that much of the controversy concerning the question had in fact arisen from a misunderstanding of the exact problem at hand; namely, that of assessing the breeding value of a sire regardless of the level of production of the cows he happened to be mated to. The above formula was developed upon the assumption that
systematic errors such as differences in the environment of the mates of sires to be compared or between the dams and daughters were negligible. Lush (1931) concluded that little accuracy was gained by increasing the number of daughters much above three or four if such systematic errors existed, since an increase in the number of daughters reduces only random errors.

Smith (1937) developed an index designed for the selection of plant lines, using Fisher's (1936) concept of "discriminant functions" to derive a linear function of observable characters as the best available guide to the genetic value of each line. Smith's method was designed for the inclusion of several characteristics, provided the relative economic importance and the genetic relations for the various characters were known. He also showed that ratios between two plant characters (the grain:straw ratio, for example) such as are commonly used by plant breeders were in general less likely to be efficient indexes than indexes based on the primary characters.
III. THE INVESTIGATION

A. Procedure

The method of mathematical deduction was most promising in the development of the present problem because many of the simplifications which must be assumed as to the genetic relations in organic populations have been elevated to the stature of empirical laws by the studies of Fisher, Wright, Haldane and innumerable others. The initial step in the present problem was the arbitrary choice of a criterion by which to measure the efficiency of an index. The criterion chosen -- that of maximum genetic progress by selection -- satisfies the practical considerations of the breeding program in that only genetic improvement is permanent. As well as permitting the construction of indexes by the well-known multiple regression technique, it permits the estimation of phenotypic or genetic changes in individual characteristics such as are caused by the selection of animals having the highest indexes.

After the factors influencing the rate of genetic change by selection were identified and their genetic bases examined, the principles of index construction were developed. Since the logical development of a theory may, solely by its appeal to reason, conceal sophistry, an index was constructed
from data collected on an actual population. The theoretical results expected from the use of the index were compared to those actually obtained in one generation not concerned in the construction of the index.

B. Theoretical Investigation

1. The factors governing rate of genetic change

The formulation of a constructive breeding program requires in its initial stages the definition of a goal. Breeders interested in improving the economic value of their product are generally interested in increasing the merit of at least several characteristics. These characteristics as a rule will differ as to their relative importance; in addition, changing economic conditions and market demands may change their relative importance even while the breeding program is in progress. Hence the aggregate genotype of an animal is defined as

\[ H = a_1 G_1 + a_2 G_2 + \ldots + a_n G_n \]  

(1)

where \( a_i \) is the relative value of a unit change in the \( i \)th character and \( G_i \) is defined for the \( i \)th character as the sum of the average effects of the individual genes which influence that character (page 11). The goal of the breeding program can thus be defined as the maximum change in the average aggregate genotype per generation. A change in the economic
situation might require a change in the relative values assigned the different characters, but this would not necessitate changing the goal of the program.

Exact determination of the relative importance of changes in the various characteristics (i.e., of the magnitude of the $a$'s) is of primary importance. Genetic interest in inducing by selection considerable change in a characteristic of little or no economic importance might be aroused, but such an adventure would entice the practical breeder only to the extent that his profits were increased. Since this question is considered in a later section, it is sufficient here to remark that the proper evaluation of the various characteristics is of fundamental importance to the economic success of the entire program.

Although the genotypes of the individual characteristics cannot be appraised exactly, the aggregate genotypic value defined above is a very real quantity, existing for each animal in the population, albeit not directly measurable. Suppose there exists, however, in addition to the aggregate genotype a second variable, an index ($I$), which can be defined exactly for each animal, and that $H$ and $I$ are normally distributed. Then if a proportion ($p$) of the animals having the highest values for this second variable are selected as parents of the next generation, the average gain in the aggregate genotypic value of the selected group over the unselected group is expected to be
\[ E(I) = \frac{Z}{p} R_{IH} R_H \]  \hspace{1cm} (2),

where \( Z \) is the height of the ordinate of the normal curve at the lowest value of \( I \) retained in the selected group and \( R_{IH} \) is the correlation between the aggregate genotypic value and the new variable \( I \).

The index can be defined mathematically as

\[ I = b_1x_1 + b_2x_2 + \cdots + b_nx_n \]  \hspace{1cm} (3),

where the \( b_i \)'s are regression coefficients chosen such that \( R_{IH} \) is a maximum, and the \( x_i \)'s refer to simple or perhaps complicated linear functions of the various characters in the selection program. That is, \( x_i \) can be a single production record, the average of several records, the performance of a group of relatives, etc., concerning the \( i \)th characteristic.

The values of the regression coefficients in (3) are established as follows: Putting \( \frac{2}{b_1} \log R_{IH} = 0 \), since \( R_{IH} \) is a maximum when \( \log R_{IH} \) is, \( n \) simultaneous equations are obtained. They are shown here in standard measure (Snedecor, 1940) because of the convenience in calculating the necessary constants in animal populations. That is, the constants can be expressed as correlation coefficients which do not require the same number of observations in each. The equations are
\[ B_1 + B_2 r_{x_1 x_2} + \ldots + B_n r_{x_1 x_n} = r_{X_1 H} \]
\[ B_1 r_{x_1 x_2} + B_2 + \ldots + B_n r_{x_2 x_n} = r_{X_2 H} \]

\[ B_1 r_{x_1 x_2} + B_2 r_{x_2 x_2} + \ldots + B_n = r_{X_n H} \]

(4)

where \( B_1 = \frac{\sigma_{X_1}}{\sigma_H} \), and the various \( r \)'s are correlation coefficients between the designated variables. The exact technique of deducing the value of the correlation coefficient involving \( X_1 \) will depend upon the particular kind of function \( X_1 \) happens to be. A number of such techniques are given as examples in a later section. It is sufficient to note here that the method of path coefficients makes possible their evaluation.

Assuming the necessary constants in (4) can be evaluated, values of the regression coefficients in (3) are found by first determining the values of the \( B \)'s as follows:

\[ B_1 = C_{11} r_{X_1 H} + C_{21} r_{X_2 H} + \ldots + C_{n1} r_{X_n H} \]

(5)

where \( C_{ij} \) is the cofactor of the \( i \)th row and \( j \)th column divided by the determinant (D) \([\text{including the sign } (-1)^{i+j}]\) where

\[ D = \begin{vmatrix}
1 & r_{x_1 x_2} & \ldots & r_{x_1 x_n} \\
r_{x_1 x_2} & 1 & \ldots & r_{x_2 x_n} \\
r_{x_1 x_3} & r_{x_2 x_3} & \ldots & 1 \\
r_{x_1 x_n} & r_{x_2 x_n} & \ldots & 1 \\
\end{vmatrix} \]

(6).
The genetic progress expected per generation can be calculated for any particular case by evaluating

$$R^2_{IH} = \sum_{i=1}^{n} B_i R_{X_i H}$$  \hspace{1cm} (7),

$$\sigma^2_H = \sum_{i=1}^{n} a_i^2 \sigma_i^2 + 2 \sum_{i<j} a_i a_j r_{G_i G_j} \sigma_i \sigma_j$$  \hspace{1cm} (8),

and $Z/p$.

2. Genetic basis for correlations between genotype and phenotype

The correlation ($r_{X_i H}$) between a component ($X_i$) of the index and the aggregate genotype ($H$) is itself a compound linear function composed of $n$ residual correlations. From equation (3) $r_{X_i H}$ can be analyzed as

$$r_{X_i H} = d_1 r_{X_i G_1} + d_2 r_{X_i G_2} + \cdots + d_n r_{X_i G_n}$$  \hspace{1cm} (9),

where $d_i = \frac{a_i \sigma_i}{\sigma_H}$. Since the constituents of the correlations between the individual components of the index and the aggregate genotypic value are the fundamental units of index construction, their use can be justified only by considering their genetic bases with reference to the present state of genetic knowledge. Methods of calculating these correlations are presented in a later section; the chief object here is the statistical interpretation of known genetic phenomena. A path coefficient diagram (fig. 5) showing the important relations of equation (9) is presented at the end of this section.
Wright (1920) explored the genetic basis for the correlation between genotype and phenotype for color pattern in guinea pigs and justified experimentally the assumption that the hereditary and nonhereditary differences affecting the phenotype were uncorrelated (page 13). By this assumption,

$$\sigma^2_{x_1} = \sigma^2_{G_1} + \sigma^2_{E_1} \tag{10}$$

where $\sigma^2_{E_1}$ now includes all sources of variance except the additively genetic. In terms of path coefficients (fig. 3),

$$1 = e_1^2 + e_1 \tag{11}$$

Then the correlation between the genotype and phenotype of the $i$th characteristic is evidently

$$r_{X_1 G_1} = \frac{e_1}{\sigma_{X_1}} = \frac{\sigma_{G_1}}{\sigma_{X_1}} \tag{12}$$

The basis upon which the accuracy of these relations depends is that the hereditary and nonhereditary differences are uncorrelated, an assumption which is justified unless particular efforts are made to give the more desirable genotypes the better environment or vice versa.

Straus (1940) commented upon the existence of correlations between the genotype ($G_j$) of one character and the phenotype ($X_i$) of another. Upon the same assumption as before,

$$r_{X_1 G_j}$$

can be analyzed as

$$r_{X_1 G_j} = e_1 r_{G_1 G_j} \tag{13}$$

where $e_1$ is the correlation between the genotypes of dif-
Fig. 3. Path coefficient diagram to illustrate the statistical relations between the genotypes for two characteristics ($G_i$ and $G_j$) of an animal and the phenotype for one of the characters ($X_1$). In this study differences in environment, epistatic interactions and dominance deviations (fig. 2) are included in $E_i$. 
ferent characteristics (referred to henceforth as a **genotypic correlation**). These relations are shown in fig. 3.

Genotypic correlations might exist in a Mendelian population because of linkage, because of selection with varying emphasis upon the different characters in isolated subdivisions of the population or because of genes with pleiotropic effects. The effect of linkage upon genotypic correlations would be transitory, arising only after a population originated from a cross of different lines and decreasing asymptotically as the linkage combinations approached equilibrium.

The effect of selection upon genotypic correlations depends upon the extent to which selection intensity varies in relatively non-interbreeding units of the population, the number of characteristics involved and the effectiveness of selection. Considering only two characteristics and the case of no pleiotropic genes, if one group of breeders were successful in increasing the expression of character A by selection but paid no attention to character B, while the rest were successful in increasing the expression of B but paid no attention to A, the genotypic correlation would be negative, increasing in magnitude as the two groups were drawn farther apart by selection. If one group of breeders sought an increase in A and a decrease in B, while the rest selected for the opposite ideal, the genotypic correlation would be more strongly negative than in the former case insofar as selection was successful. Simultaneous selection for
increased expression in both $A$ and $B$ by one group and decreased expression in both characters by the other group would engender a positive correlation between $G_A$ and $G_B$. Selection in a positive direction for both characteristics, with varying emphasis on the two in different herds, would tend to generate a negative genotypic correlation -- albeit a curvilinear rather than a linear one. The Shorthorn breed is an eminent example of the latter case; the degree of emphasis placed upon beef conformation and milk production varies widely, but selection is directed toward increased performance in both characters. A second example may be seen in dairy cattle where some breeders select chiefly for type, paying little attention to production, while other breeders select rigidly for production with slight regard for type. The example, of course, is more extreme than the actual case in that there are all gradations of emphasis placed upon the importance of the two characters if all herds in a dairy breed are considered as the population.

The transient nature of herds in time and the exchange of breeding animals between herds usually would prevent the isolation necessary for the development of noticeable negative genotypic correlations. If a difference in ideals persisted for a long period of time as in the Shorthorn example, the exchange of animals between herds would become limited to herds having similar ideals.

Pleiotropic genes -- genes which affect more than
one character -- are a possible source of genotypic correlations. Dobzhansky (1937) suggests "that most, and possibly all, genes have manifold effects... In general, there is no conclusive evidence to show that genes have a circumscribed province including only one class of characters as physiological reactions; our understanding of the dynamics of individual development is still in its early infancy." Hence it appears of little consequence to speculate on the nature of their origin or mode of producing their manifold effects. Any gene which has an effect upon the phenotypes of two characters, regardless of whether the effect is "direct" or "indirect" or whether it may affect only one character in some gene combinations, contributes some of the heritable variance associated with each character and hence some of the covariance common to their genotypes. Neither does there seem to be any scientific basis for postulation as to the direction of influence a pleiotropic gene may have on the respective characters it affects.

The statistical consequences of the effects of simultaneous selection for characteristics determined in part or wholly by pleiotropic genes can be determined in a general way. In a population where only two characteristics are being selected for, genes can be grouped into four classes according to their effects: (1) They act on character A only; (2) they act on character B only; or (3) they have the effect of increasing or decreasing both A and B; or (4) they have opposite
effects upon $A$ and $B$. By assuming equal selection coefficients for genes in classes (1) and (2), a selection coefficient twice as large for those in class (3) and a coefficient for those in class (4) of zero, the effects of selection can be studied for an isolated case.

The frequency of desirable genes in classes (1) and (2) will increase, the rate depending upon the selection coefficient as well as upon the gene frequency and degree of dominance. The frequency of desirable genes in class (3) will increase relatively more rapidly than those in classes (1) and (2) and in addition will maintain a higher frequency at equilibrium if the mutation rates in the three classes are at all equal. Any change in the frequency of genes in class (4) will occur by chance.

The contribution to the genotypic variance for any characteristic from a particular pair of genes is $2q(1-q)\gamma^2$ (Wright, 1931), where $q$ is the frequency of one allele and $\gamma$ is the average effect of the gene. If a gene affects two characters, its contribution to the variance of $A$ and $B$ is $2q(1-q)\gamma_A^2$ and $2q(1-q)\gamma_B^2$, respectively, whereas its contribution to the covariance is $2q(1-q)\gamma_A \gamma_B$. When the frequency of a gene is very high or very low, its contribution to the genetic variance or covariance is relatively small as compared to the contribution of a gene whose frequency is between .3 and .7. Hence, the ultimate effect of selection upon the genetic variance from genes in classes (1) and (2) and upon the variance
and covariance from genes in class (3) is to make it decreasingly small. Even if the selection coefficients for genes in class (4) are not exactly zero, they will be considerably smaller than the coefficients of the genes in the other classes; hence, the genes themselves will reach equilibrium at a lower frequency, and their contribution to the covariance 

\[-2q(1-q)\gamma_{A_B}\]

will be relatively greater.

The conclusion is inevitable that negative genotypic correlations generally prevail under conditions where: (1) Many of the genes have pleiotropic effects; (2) the various effects of a pleiotropic gene are random (as likely to be positive as negative for a character regardless of its other effects); and (3) effective selection has swept the frequencies of favorable genes with single effects to high levels. The limit approached by a genotypic correlation under the above conditions is -1 if only two characters are being selected for. In a population where selection intensity is equal for \(k\) characters, the limit of the genotypic correlation is \(-\frac{1}{k-1}\) if the pleiotropic effects of genes are random as regards the characters they affect.

The relations existing between the genotypes for several characters and between the genotypes and phenotypes are shown in fig. 4. The phenotypic correlation can be analyzed as

\[r_{X_iX_j} = e_{i} e_{j} G_i G_j + e_{i} e_{j} E_i E_j\]  

(14).
Fig. 4. Path coefficient diagram to illustrate the analysis of observed phenotypic correlations into genotypic and environmental portions. The G's represent genotypes for different characters present in each animal.
The random environmental factors to which an animal is exposed or a correlation between the dominance or epistatic deviations of pleiotropic genes may affect two characteristics possessed by him, either in the same direction or in opposite directions, thus contributing to the observed correlation between the characters. In view of the relatively low values of \( R_1 \) found for quantitative characters in livestock -- except for skeletal dimensions, etc. -- it is evident that the genotypic contribution to the correlation between phenotypes is small except in cases where the genotypic correlation and the heritability of the two characters is unusually high. The conclusion can generally be drawn that the correlation between two phenotypes of an animal is but a poor indicator of the genotypic correlation between those two characters.

3. Principles of index construction

The factors which control the rate of genetic change have been identified, and the genetic basis for various relations between genotypes and phenotypes have been investigated. This constitutes the material necessary for the development of an index, the use of which will permit the attainment of maximum genetic change. The principles to be followed in the construction of an index require that each of the factors identified in equation (2) be as large as possible.

a. The selection differential. Values of \( Z \) for
Fig. 5. Path coefficient diagram to illustrate the compound nature of the correlation between the aggregate genotype (H) and the phenotype for the $i$th character.
given values of $p$ are recorded in tables of the normal probability integral, available in many statistical textbooks. Tabulated values of the ratio $Z/p$ may be found in Table II of Pearson's "Tables for Statisticians and Biometricians, Part II."

The index determines which animals shall fall within the selected group, but otherwise the proportion saved as parents of the next generation is an independent quantity determined by the rate of reproduction, longevity, etc., of the species. The breeder may increase the selection differential somewhat by increasing fertility, reducing death losses and other managerial practices.

Probably the present methods of disease control are the sources of greatest loss in lowering the selection differential. For example, almost every breeder of cattle has experienced the loss of some of his best animals when they reacted to the tuberculin test, a test that indicates only that the reactor harbors the organism and not that the animal is "susceptible" to its ravages. The sounder biological practice would be to prize more highly those animals which contact the organism but resist its effects, especially since disease resistance has been found to have a genetic basis in so many instances. It is certain that present methods of disease control must demand their expensive toll among the best breeding stock in the country as long as the causative organisms exist or until preventive methods of vaccination are dis-
covered.

b. The aggregate genotypic variability. The standard deviation of the aggregate genotypes ($q_H$) is fixed, once the goal of the breeding program is defined, except for modification due to changes in the heritable portion of the variance for the various characters in the program. This is suggestive of Fisher's (1930) Fundamental Theorem of Natural Selection: "The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time." Rewriting equation (3) as

$$
q_H^2 = a_1^2 q_1^2 + a_2^2 q_2^2 + \cdots + 2a_1 a_2 q_1 q_2 r_{12} + \cdots
$$

the roles of the individual variances are evident. In the initial stages of a breeding program the frequency of desirable genes may be relatively low, and selection will cause a temporary increase in the variance, thus increasing the rate of progress. However, long-continued selection, if successful, must eventually lead to higher gene frequencies or, if the phenotypically most desirable is a genetic intermediate, to smaller average effects of the genes, and thus to less additively genetic variability so that additional progress becomes increasingly difficult.

Very high negative genotypic correlations, while in no way affecting the genotypic variances for the individual characters, decrease the aggregate genotypic variance sharply.
This theoretically could lead to the paradoxical situation where selection was powerless, although genetic variability for each character still existed, because gains in one characteristic were exactly cancelled by losses in another.

In beef cattle, sheep and swine much importance is attached to such characteristics as cut-out value, quality of carcass, dressing percent, etc., which are not directly measurable in animals saved for breeding. Such characters can be included in $\mathbf{H}$ by crediting an animal with the records of relatives and/or by measuring characters of the live animal with which they are correlated genotypically; in fact, $\mathbf{H}$ is completely defined only when every characteristic which contributes to the net profitability of the animal is included.

**c. The correlation between the aggregate genotype and the index.** The technique employed in determining the values of the regression coefficients ($b$'s) in the index permits the most efficient use of the information available in estimating the aggregate genotypic value of each animal for which an index is calculated. The size of $R_{IH}$ is limited by the amount of information concerning the genotypes for the specific characteristics of the breeding program that is contained in the components of the index (the $x$'s). As defined, the $x$'s have only one restriction — that correlations between them and the genotypes for the different characteristics can be calculated. The method of path coefficients is particularly well adapted to the analysis of correlations between geno-
types and complicated systems of variables. Thus critical use of information from relatives, lifetime averages and characteristics having no economic importance themselves may serve to increase the correlation between the aggregate genotype and the index.

Correction factors for age and for number of times milked daily are in general use in comparing production records of dairy cattle. They have not been used extensively in types of livestock other than dairy cattle, probably because widespread interest is only now being aroused in measuring the performance of individual animals. The individual breeder, having the opportunity of observing his animals closely and being aware of some of the environmental influences to which his animals are subjected, might be able to employ correction factors to much greater advantage than is often done at present. The correction should be applied with the view of making the corrected record what the actual record would have been had the particular influence under consideration been the standard one. Disadvantages of correction factors are: (1) Difficulty in measuring accurately the effect of a particular extraneous influence and (2) some risk that the individual breeder might use them with unconscious prejudice or favoritism.

The primary information concerning the genotype of an animal for any particular characteristic comes from two
sources:

(1) His own performance for that characteristic,
    (a) one record
    (b) n records

(2) Performance of his relatives for that characteristic,
    (a) parents
    (b) collateral relatives
    (c) offspring

Information which may be considered as secondary comes from other characteristics because of the correlation between genotypes for different characters.

The correlation between genotype and phenotype for the same characteristic was given in equation (12) where the phenotype was considered as one record. In many types of livestock the phenotype is repeatable, in the sense that it is given new expression at more or less definite intervals during an animal's lifetime. Since temporary environmental influences which affect only single records tend to cancel each other, increasing the number of records causes an increase in the correlation between the genotype and the average of the records. The deceptive effects of dominance, epistasis or permanent environmental influences which extend over the entire period that the records are taken are not reduced by increasing the number of records. The correlation
between the genotype and the average of \( n \) records is

\[
r_{G_iX_i} = g_1 \sqrt{\frac{n}{1 + (n-1)r_{XY}}} \tag{15}
\]

where \( X \) represents a single record and \( g_1 = \frac{\partial G_i}{\partial X_1} \). The derivation of this formula requires the assumptions that all records are equally correlated and have equal variances. Actual experience with livestock records has shown that these assumptions are valid, since the variances of records are reasonably constant and the correlations between records are similar (Berry, 1939). The statistical relations between the genotype and the average of a number of records are shown in fig. 6.

The formula for the correlation between \( X_i \) where

- \( X_i \) is the average of a number of records or measurements of the \( i \)th characteristic and the genotype for another character (\( G_j \)) is

\[
r_{X_iG_j} = r_{G_iG_j} g_1 \sqrt{\frac{n}{1 + (n-1)r_{XY}}} \tag{16}
\]

and may be considered a measure of the secondary information obtained from \( X_i \).

The information about an animal's genotype which can be obtained from the performance of his parents or more remote ancestors is limited because: (1) The parents are usually a selected group themselves; (2) their genotypes are not known exactly; and (3) the sampling nature of inheritance
Fig. 6. Path coefficient diagram to illustrate the compound nature of the correlations between the genotype and average of $n$ records for the same character ($r_{G_1X_1}$) and between the average of $n$ records and the genotype of another characteristic ($r_{G_jX'_1}$).
the exact prediction of the genotype of an offspring from heterozygous parents. Hence, information obtained from parents or other ancestors is of most importance when little or no information is available concerning the animal's own performance.

Wright (1921b) investigated the biometric relations between parents and offspring, showing that the direct influence of the parent (path coefficient $ab$ in fig. 7) upon the offspring was $1/2 \frac{1 + f'}{1 + f}$, where $f'$ and $f$ represent the inbreeding coefficients (Wright, 1922) for the parents and offspring, respectively. The correlation between the record of one parent ($X_i$) and the genotype of the offspring ($G_i$) for the same characteristic is

$$r_{G_iX_i} = abg_i(1 + m)$$  \hspace{1cm} (17),

where $m$ is the coefficient of relationship (Wright, 1922) between mates. The above formula applies specifically to populations not under selection where consanguineous mating or random mating ($m=0$) exists. The secondary information contributed by $X_i$ concerning the genotype for the $i$th characteristic of the offspring is measured by

$$r_{G_iX_j} = abg_i(1+m)r_{G_iG_j}$$  \hspace{1cm} (18).

The information concerning an animal's genotypes gained from ancestors more distant than grandparents is small, although this depends somewhat upon how well the genotypes of
Fig. 7. Path coefficient diagram to illustrate the biometric relations between a parent's phenotype for the $i$th character and the genotype for the $i$th and $j$th characters of an offspring.
intervening ancestors are known. A general formula which includes information from a number of ancestors simultaneously would be so complex without a number of simplifying assumptions as to have little practical value.

The formula for the correlation between the performance of collateral relatives and the genotype of an animal can be simplified considerably by grouping relatives according to their relationships to the animal in question, i.e., as full sibs, as half sibs, etc. The relations between the genotype of an animal and the average performance of several animals having the same genetic relationship to him \((q)\) are illustrated in Fig. 8. The correlation between their average performance and the animal's genotype is

\[
\rho_{X_1} = q\sqrt{\frac{n}{1 + (n-1)\rho_{YY}}} \tag{19}
\]

An optional method of handling the information contributed by collateral relatives is offered by the progeny test.

Information obtained from an animal's offspring has long been used as a measure of his breeding value, particularly for characteristics that are not expressed by one sex. The principal limitations to such information which arise in actual practice are: (1) A long time is required to obtain information on offspring, (2) progeny of a given animal may be exposed to the same seasonal and managerial environment and (3) only the records of selected progeny are obtained.
Fig. 8. Path coefficient diagram to illustrate the biometric relations between the average performance of a group of relatives ($X_i$) and the genotype of another, all of which have the same coefficient of relationship ($q$).
The relations existing between the genotype of an animal and the records of his progeny are shown in fig. 9. The correlation between his genotype and the average performance of his progeny for the same character is

\[ r_{X_1} = \frac{abg(1+m)\sqrt{\frac{\beta}{1 + (n-1)p_{XY}}}}{1 + (n-1)p_{XY}} \]  

(20)

This formula is derived with the assumption that the records of offspring are equally correlated and have equal variances.

The preceding examples of methods of calculating the correlation between the genotype of an animal and a more or less complex function of its own performance or the performance of relatives are sufficient to indicate the usefulness of such methods in index construction. Usually the more complicated correlations can be expressed in terms of their more fundamental biological constituents which are: (1) Relationship coefficients, which can be calculated by Wright's (1922) method; (2) the heritable fraction of the variance, which can be calculated by several methods enumerated by Lush (1941); and (3) the phenotypic correlations between the individual records, which can be calculated by well-known interclass or intraclass correlation methods. Hence, these functions may be used as components of an animal's index if their use increases the correlation between his index and aggregate genotype. A path coefficient diagram in fig. 10 shows the relations between \( H \) and \( I \) schematically.
Fig. 9. Path coefficient diagram to illustrate the biometric relations between the average record of \( n \) offspring \( (X_1) \) and the genotype of a common parent \( (G_{SIRE}) \).
Fig. 10. Path coefficient diagram to illustrate the relations between the aggregate genotypic value (H) and an index (I) constructed so that \( R_{IH} \) is a maximum.
4. Interpretation of theoretical deductions

Some of the principles underlying the index developed herein have been advocated and practiced in livestock improvement for countless generations. Darwin, for instance, noted that primitive peoples preserved their best animals during periods of famine even to the extent of sacrificing the older members of the tribe. Varro recognized the merit of the progeny test as an indicator of the sire's breeding value 2,000 years ago. Pedigree selection is widely practiced by breeders of purebred animals, particularly of young animals. Pearl used family selection as a method of improving egg production in the classic Maine experiments. Goodale (1937) employed a system of "selective breeding" in selecting for increased weight in mice, evidently meaning by this that a combination of the progeny test, family performance and individual merit was the criterion of selection.

A logical method of constructing an index has been proposed which permits the incorporation of the information upon which each system is based into a single quantitative value for each animal in the herd, the different bits of information being automatically weighted according to their relative value in estimating the actual breeding value of the animal. The index is so devised as to utilize information for several characteristics and for traits which cannot be measured in the breeding animals themselves.
The principles of the index have made it apparent that some of the practices and generally accepted theories concerned with selection are not wholly correct. In addition they appear to make possible the solution of problems which have arisen previously but have not as yet been treated adequately. The degree of success expected by following a stated breeding program can be predicted a priori, provided the biological constants necessary for the construction of the index are known. This permits the comparison of various procedures in the light of the expected progress attending each.

a. Methods of selection. Some confusion has accompanied the dissemination of knowledge concerning the selection experiments carried on by research institutions as to the most efficient method of selection, particularly since the very nature of the problem in many such experiments appeared to require selection for only one characteristic. The effects of selection for unimportant traits, fads and fancy points in lowering the selection differential for economic characteristics has been frequent cause for comment. The natural sequence of events has led to the popular impression that for selection to be most effective it must be directed at only one character at a time.

The method of "culling levels" for each characteristic, when selection is being practiced simultaneously for
several characters, has also been proposed. A more or less
definite level, determined largely by the amount of culling
that can be practiced and by the number of characteristics
in the selection program, is set for each trait, animals be-
low the minimum level in any characteristic being discarded
regardless of their merit in other characters.

The genetic progress expected under these two
methods of selection can be compared to that expected by us-
ing as the criterion of selection an index constructed to
produce maximum genetic progress. Assume \( n \) characters are
equally important, equally hereditary and uncorrelated geno-
typically or phenotypically. The genetic progress expected
in one generation is

\[
E(H) = E(G_1) + E(G_2) + \cdots + E(G_n).
\]

The progress expected by the first method, that of selecting
for one characteristic at a time is

\[
E_1(H) = E_1(G_1) = \frac{Z_1}{p} \frac{\sigma^2_{G_1}}{\sigma^2_{X_1}}.
\]

That expected by the method of culling levels is

\[
E_2(H) = nE_2(G_1) = n \frac{Z_2}{\sqrt{p}} \frac{\sigma^2_{G_1}}{\sigma^2_{X_1}}.
\]

The progress expected by the index method where every char-
acter is given proper weight is

\[
E_3(H) = \sqrt{n} E_1(G_1) = \sqrt{n} \frac{Z_1}{p} \frac{\sigma^2_{G_1}}{\sigma^2_{X_1}}.
\]
Then

\[ E_3(H) > E_2(H) > E_1(H), \text{ for } n > 1. \]

In particular,

\[ \frac{E_3(H)}{E_1(H)} = \sqrt{n}. \]

\( E_3(H) \) decreases as \( p \), the proportion saved, increases, and \( E_2(H) \) increases as \( n \), the number of characters, increases, but is always greater than unity. Curves of \( E_1(H) \), \( E_2(H) \) and \( E_3(H) \) are shown for values of \( p \) from 1.00 to .01 and \( n = 4 \) in fig. 11.

When the relative importance or the heritability of the characters differs, the first method, that of selecting for only one character at a time, compares more favorably with the other methods than is indicated above, if only the most important character is selected for and it is as heritable as the other characters. The progress by the first method might actually surpass that made by using culling levels, because too much attention might be paid to unimportant or poorly heritable characters by the second method. The third method, that of index selection, has the important property of weighting the amount of attention given each character so that genetic progress is a maximum for a given selection intensity; thus the other methods could not surpass it in efficiency. The effect of positive genotypic correlations is to increase the progress by any of the methods, but
Fig. 11. The genetic progress expected by different methods of selection.
it is not increased equally by all methods. The effect of negative genotypic correlations is to decrease the total genetic progress, although progress in any one characteristic could be maintained.

The advantage indicated for the selection of several characters over that of selecting for one character holds only so long as the traits have real economic importance. Critical survey of the economic importance of any character should be made before including it in the selection program. When selection is directed toward several characters that are equally heritable and are considered as equally important, the progress for any one character is only \( \frac{1}{\sqrt{n}} \) times as great as if selection were solely for that trait. Thus, if some characters having no economic value were selected for, or if some characters having economic value were ignored, full advantage would not be taken of the opportunity afforded by selection.

**b. Genetic progress by indirect selection.** The ineffectiveness of selection is emphasized in the breeder's mind by the regression of the offspring of individually superior animals toward the breed average. The breeder is aware of the brevity of his lifetime in terms of livestock generations, realizing that it is all too short for much to be accomplished. The individual merit of animals is difficult to measure accurately for many characters or can be measured only after maturity is reached and often requires the time-
consuming practice of recording and analyzing data. These factors have led to the emphasis of characteristics which are correlated with economically important characteristics, many of which are associated with the terms "type" and "conformation." There are two objections to such indirect attempts to create genetic improvement in economically important characteristics: (1) The selection differential is lowered for each additional character considered, and (2) a study of the genetic basis for the correlation between characters indicates that, except under very unusual conditions, the use of correlated characters in selecting for another character is likely to result in decreasing progress.

The average of the selected group for a character such as fat production in dairy cattle is increased by selecting directly for a positively correlated character such as type. Casual observation has led to the belief that the gain in average fat production in the selected group was, in part at least, genetic. But this is not necessarily true. The genetic gain in fat production depends altogether on the genotypic correlation between the two characters if only type has been selected for directly. The many accidental environmental factors to which animals are exposed can be regarded as random with respect to any one character in that they increase or decrease the phenotype as compared to the genotype for that character. But an environmental factor which affects
one character of the animal probably affects other characters possessed by him; thus the environmental deviations in characters of the same animal have reason to be correlated. Unless there is a positive genotypic correlation, the increased fat production due to direct selection for type is in reality due only to the favorable environmental factors to which the selected animals were exposed. Their genetic average is expected to be the same as that of the unselected group as far as fat production is concerned.

The point in question can be illustrated by an example. Let

\[ G_{X_1}^2 = G_{X_2}^2, \]

\[ G_{G_1}^2 = G_{G_2}^2, \]

\[ r_{X_1X_2} \neq 0 \quad \text{and} \quad r_{G_1G_2} = 0. \]

Assume that \( X_2 \) has no economic value of itself alone. Then

\[ H = a_1G_1 + a_2G_2 = G_1. \]

Four types of selection programs deserve study:

1. Indirect selection for \( X_1 \) by selecting directly for \( X_2 \), so that

\[ I_1 = X_2; \]

2. Equal attention given to \( X_1 \) and \( X_2 \), so that

\[ I_2 = X_1 + X_2; \]
(3) direct selection for \( X_1 \), so that
\[
I_3 = X_1,
\]
and

(4) direct selection for an index designed to make genetic
progress a maximum, so that
\[
I_4 = b_1 X_1 + b_2 X_2 = X_1 - r X_1 X_2.
\]

Then the genetic progress expected by the four methods is
\[
E_1(H) = 0,
\]
\[
E_2(H) = \frac{2}{p} \frac{\partial^2 G_1}{\partial X} \sqrt{\frac{1}{2 \left[ 1 + r X_1 X_2 \right]}},
\]
\[
E_3(H) = \frac{2}{p} \frac{\partial^2 G_1}{\partial X},
\]
\[
E_4(H) = \frac{2}{p} \frac{\partial^2 G_1}{\partial X} \sqrt{\frac{1}{1 - r^2 X_1 X_2}},
\]

whence
\[
E_4(H) > E_3(H) > E_2(H) > E_1(H).
\]

General conclusions can be drawn from this illustration subject to the assumption that \( \Sigma G_1 G_2 \) is zero as follows:

(1) Indirect selection alone is completely ineffective
in causing genetic change in a correlated character.

(2) The generally approved practice of giving at least
some credit for a positively correlated character
lessens the genotypic progress.

(3) An index considering both characters and constructed
in accordance with the principles suggested in this study satisfies the criterion of maximum progress.

Perhaps the most striking feature of the comparison lies in the use of the character $X_2$ by the fourth method. With a positive phenotypic correlation, high values of $X_2$ detract from the animal's index; with a negative phenotypic correlation, high values of $X_2$ enhance the index, whereas the opposite is the generally approved practice. The logic for this procedure has been mentioned; namely, values of $X_2$, attributed in part to exceptionally good or exceptionally poor environment, serve as an indicator of the kind of environmental deviations which have affected the economically important characteristic.

While these examples are admittedly crude in view of the tremendously complex situations they represent, they do seem to illustrate the folly of indirect selection unless something is known of the genotypic correlations underlying the expression of characters. The role of the character which has no economic importance in constructive breeding is thus to serve as an indicator of (1) the genotype of an important character when a genotypic correlation exists and (2) the random environmental factors which affect economically important characters.
C. Application of the Index to an Actual Population

1. Source and adjustment of data

The data used in the construction of an index were taken on 12 lines of Poland-China swine at the Iowa Agricultural Experiment Station. These lines are maintained as a cooperative breeding project by the Iowa Station and the Regional Swine Breeding Laboratory, their history being given by Whatley (1939). The data used in constructing the index include pigs born in the 12 lines from the fall of 1937 to the spring of 1940, inclusive. The data on the pigs born in the fall of 1940 were used as a test of the index. A brief summary (table 1) shows the quantity of data by seasons.

Selection has been practiced for three characteristics, (1) growth rate as measured by weight at 180 days of age, (2) conformation as measured by a numerical score and (3) productivity as measured by number and weight of pigs per litter at various ages, more or less systematically since the lines were founded. These characteristics were weighted according to their economic importance as determined by Lush (1938a, 1940) and an index was computed for each animal since 1938. Except for the elimination of animals because of injury and disease, selections of breeding stock have been made upon the basis of this index. More recently two additional factors have been added to the index, the performance of litter mates
in 180-day weight and in score being considered. Lush (1938b) studied the theoretical basis for the relative amounts of attention which the performance of relatives should receive.

Table 1. Quantity of Data Listed by Seasons.

<table>
<thead>
<tr>
<th>Season</th>
<th>No. of lines</th>
<th>No. of litters</th>
<th>No. of pigs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1937 fall</td>
<td>1</td>
<td>12</td>
<td>61</td>
</tr>
<tr>
<td>1938 spring</td>
<td>1</td>
<td>77</td>
<td>391</td>
</tr>
<tr>
<td>1938 fall</td>
<td>1</td>
<td>20</td>
<td>97</td>
</tr>
<tr>
<td>1939 spring</td>
<td>12</td>
<td>94</td>
<td>454</td>
</tr>
<tr>
<td>1939 fall</td>
<td>7</td>
<td>28</td>
<td>158</td>
</tr>
<tr>
<td>1940 spring</td>
<td>11</td>
<td>83</td>
<td>385</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>314</td>
<td>1,546</td>
</tr>
</tbody>
</table>

The accumulation of data during the last several years was thought to justify the construction of an index based upon the genetic constitution underlying the determination of those characteristics considered to be of economic importance in swine production.

Litters received substantially the same kind of treatment during the suckling period. Four to eight litters were usually placed in pasture lots at about 10 days of age, remaining there until they were weaned at 56 days. Four pigs from the largest litters were usually placed on "Record of Performance" tests at weaning time, being fed a standard diet
in small pens with concrete floors where accurate records could be obtained on their feed consumption. The rest of the pigs were distributed at random in various treatments, all treatments being reasonably good ones and differing little in results. All of the pigs were weighed as nearly as possible at 180 days of age, an occasional pig being weighed at more than 4 days away from this age. Their weights were corrected to 180 days of age by the equation

\[
180\text{-day weight} = \frac{120}{\text{actual age} - 60} \times \text{actual weight}
\]

which was suggested by Whatley (1937). No corrections were made for treatment differences except for pigs on Record of Performance. In this case the average difference between the pigs on test and their litter mates distributed in other treatments was added to the weight of each pig on test.

With few exceptions the pigs, when weighing between 215 and 235 pounds, were scored by a group of at least 4 judges. The judges independently assigned values between 0 and 9 to each of six traits of the animal, the score being the sum of the six values. The traits on the score card were:

1. Vigor, health and thriftiness
2. Quality
3. Length of body
4. Details of conformation
5. Animal as a whole
6. Grade
The measure of productivity used in this study was suggested by Lush and Molln (1941) who studied the repeatability of number of pigs per litter and of weight per litter at a number of Stations cooperating with the Regional Swine Breeding Laboratory. The characteristics used in calculating productivity were weighted so that one standard deviation in each had approximately the same influence in determining differences in productivity. The characteristics considered and their weights are shown in table 2.

Age correction factors found by Lush and Molln to eliminate most of the age effect upon productivity, shown in table 3, were used.

2. Constants necessary for the construction of the index

Selection in the swine breeding project is being directed chiefly toward the improvement of three characteristics; namely (1) weight at 180 days of age, (2) score at 225 pounds and (3) productivity, the method of measuring these characteristics having been described. Lush (1938a, 1940) studied the relative economic importance of a change of one standard deviation in each of these characters, arriving at values of $1/3$, $1$ and $2$ respectively for the units in which each is measured. In accordance with these values, the aggregate genotypic value of an animal may be defined as

$$R = \frac{1}{3} G_w + G_s + 2G_p.$$
Table 2. The Components and Their Weights Used in Calculating Productivity.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of pigs born</td>
<td>1</td>
</tr>
<tr>
<td>Number of pigs alive at 21 days</td>
<td>1</td>
</tr>
<tr>
<td>Number of pigs alive at 56 days</td>
<td>1</td>
</tr>
<tr>
<td>Weight of litter at 21 days</td>
<td>1/10</td>
</tr>
<tr>
<td>Weight of litter at 56 days</td>
<td>1/30</td>
</tr>
</tbody>
</table>

Table 3. Age Correction Factors Used in Calculating Productivity.

<table>
<thead>
<tr>
<th>Age (months)</th>
<th>Correction factor</th>
<th>Age (months)</th>
<th>Correction factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0</td>
<td>36</td>
<td>-6.3</td>
</tr>
<tr>
<td>16</td>
<td>-7.3</td>
<td>42</td>
<td>-5.3</td>
</tr>
<tr>
<td>24</td>
<td>-14.0</td>
<td>48</td>
<td>-5.3</td>
</tr>
<tr>
<td>30</td>
<td>-7.6</td>
<td>54</td>
<td>-3.0</td>
</tr>
</tbody>
</table>
At the time selections are made for the yearling gilts and boars, the weights and scores are available on most of the pigs for that season. Their own productivities are, of course, unknown until after they have actually produced litters themselves. After animals are once placed in the breeding herd, later selections among them are based chiefly on their own productivity, freedom from disease and general soundness. A selection index which conforms to this system of selection and permits utilizing the information available when culling must be done is

\[ I = b_w W + b_s S + b_{\bar{W}} \bar{W} + b_{\bar{S}} \bar{S} + b_P P, \]

where \( W \) = individual 180-day weight,
\( S \) = individual 225-pound score,
\( \bar{W} \) = average 180-day weight of the litter,
\( \bar{S} \) = average 225-pound score of the litter,
\( P \) = productivity of the dam

and the \( b \)'s represent the weight given to each characteristic, respectively.

The actual values of the \( b \)'s were found by calculating the constants required by equations (6) and (9) from the data collected on the swine herd from the fall of 1937 through the spring of 1940. These constants, grouped according to their biological significance, are:

(a) Phenotypic constants

1. \( r_{ws} \), the correlation between the weight and score
of the same pig

2. \( r_{wp} \), the correlation between the weight and
dam's productivity

3. \( r_{sp} \), the correlation between the score and dam's
productivity

4. \( r_{ww} \), the correlation between the weights of litter
mates

5. \( r_{ss} \), the correlation between the scores of litter
mates

6. \( r_{ws} \), the correlation between the weight and score
of litter mates

7. \( \sigma_w \), the standard deviation of weight

8. \( \sigma_s \), the standard deviation of score

9. \( \sigma_p \), the standard deviation of productivity

(b) Genotypic constants

1. \( g_w^2 \), the heritability of weight

2. \( g_s^2 \), the heritability of score

3. \( g_p^2 \), the heritability of productivity

4. \( r_{G_wG_s} \), the correlation between the genotypes
for score and weight

5. \( r_{G_wG_p} \), the correlation between the genotypes
for weight and productivity

6. \( r_{G_sG_p} \), the correlation between the genotypes
for score and productivity
In order to avoid errors in dropped data

contrary to what place, regardless of their statement of

reason or some consequence of others, they were taken to the

construction of the index required the multiplication of

states appear not to have received previous attention. Since

considerable attention, as methods of calculating these con-

The problem of estimating genotype correlations is given.

the various constraints will be considered in the next section.

questions of interest. The seasonal estimates for

attained by analyses of verance.

these differences between dating for each season independently

samples. The calculations were obtained next on the

genetic differences between seasons were excluded from the

genetic differences between times and the environmental or

seasons are made some differences are not followed in the same season, the

since all these times are being measured and set.

on farm for the estimation of genotype constraints

phenotypic constraints and of the temperature of the

of estimates of verance and covariance. For the estimation of

under consideration, the methods generally used were those

verred constraints apply according to the particular constraints

Although the methods of statistical estimation
4. Results of the analysis

a. Phenotypic constants. The variances for weight, score and productivity were calculated for each season's data, the line differences being eliminated by analysis of variance. The results of the analysis are shown in Table 4. The seasonal variances for both weight and score vary considerably in different seasons and generally in the same direction, a fact which perhaps may be explained by the variability in treatments which existed in different seasons. For example, quite a few of the late litters were attacked rather severely by influenza in the fall of 1938, whereas the early litters escaped. The quantitative measurement of the effects of such environmental factors is extremely difficult, although such

Table 4. The Mean Square Intra-Line Variances for Weight, Score and Productivity by Seasons.

<table>
<thead>
<tr>
<th>Season</th>
<th>D/f</th>
<th>Weight</th>
<th>Score</th>
<th>D/f</th>
<th>Productivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1937 F</td>
<td>60</td>
<td>717.5</td>
<td>25.2</td>
<td>11</td>
<td>93.2</td>
</tr>
<tr>
<td>1938 S</td>
<td>390</td>
<td>1255.1</td>
<td>32.8</td>
<td>77</td>
<td>95.2</td>
</tr>
<tr>
<td>1938 F</td>
<td>96</td>
<td>1304.6</td>
<td>28.9</td>
<td>19</td>
<td>95.7</td>
</tr>
<tr>
<td>1939 S</td>
<td>442</td>
<td>1033.0</td>
<td>22.6</td>
<td>82</td>
<td>96.9</td>
</tr>
<tr>
<td>1939 F</td>
<td>151</td>
<td>685.4</td>
<td>17.2</td>
<td>21</td>
<td>129.7</td>
</tr>
<tr>
<td>1940 S</td>
<td>374</td>
<td>852.0</td>
<td>13.2</td>
<td>72</td>
<td>80.6</td>
</tr>
<tr>
<td>All seasons</td>
<td>1513</td>
<td>1015.3</td>
<td>22.88</td>
<td>232</td>
<td>94.5</td>
</tr>
</tbody>
</table>
variable factors might explain the difference in seasonal variances.

The seasonal variances for score and weight were subjected to the $X^2$ test for homogeneity (Bartlett, 1936), which is approximately

$$X^2 = N \log_e \bar{V} - \sum_{i=1}^{k} n_i \log_e V_i,$$

where $n_1$ is the degrees of freedom for the variance ($V_1$) of the $i$th season, $N = n_1 + n_2 + \cdots + n_k$,

$$\bar{V} = \frac{n_1 V_1 + n_2 V_2 + \cdots + n_k V_k}{n_1 + n_2 + \cdots + n_k}$$

and the $X^2$ has $k-1$ degrees of freedom. $X^2$ for the variances of weight and score are 31.8 and 85.6, respectively, each of which is beyond the .01 level of probability. Although the seasonal variances are influenced rather markedly by factors other than sampling errors, they were averaged as indicated in the $X^2$ test for use in the index. Since the relative size of the standard deviations is the important variable, as far as the relative size of the regression coefficients is concerned, and this changed little from season to season, it seems unlikely that this source of error would ever be serious.

The variance for productivity was much more uniform than for the other two characteristics, $X^2$ being only 2.05 for the 6 seasons. Variation as small as that observed in the seasonal variances for productivity could easily be due to
There was a strong tendency for the seasonal covariances between score and weight to vary in the same direction as the variances of score and weight, again indicating that the two characteristics are affected by a group of common causes which are more variable in some seasons than others. The seasonal correlation coefficients between the weight and score of individual pigs (table 6) were calculated and averaged by Fisher's (1938) "z" transformation as follows:

\[ z_i = \frac{1}{2} \left[ \log_e (1 + r_i) - \log_e (1 - r_i) \right], \]
\[ w_i = n_i - 2, \]
\[ z = \frac{w_1 z_1 + w_2 z_2 + \ldots + w_6 z_6}{w_1 + w_2 + \ldots + w_6}, \]

where \( r_i \) and \( n_i \) represent the correlation coefficient and degrees of freedom, respectively, for the \( i \)th season. This transformation has the effect of reducing the average correlation somewhat as compared to the correlation obtained by averaging the seasonal covariances and variances. The method used, however, is less subject to bias as the seasonal variances for score and weight differ significantly from season to season.

The correlation coefficients between the weights of litter mates and between the scores of litter mates were conveniently estimated from intraclass correlations. Because of the differences in the seasonal variances, the seasonal
Correlations were calculated individually as follows:

\[ B = \text{mean square within litters}, \]
\[ B + \frac{\overline{X}}{A} = \text{mean square between litters}, \]
\[ r = \frac{A}{A + B} \]

where \( k \) represents the number of pigs per litter and \( \overline{X} = \frac{2k}{n} \), the slight correction \( \left( \frac{\hat{e}^2}{2k} \right) \) necessary where \( k \) varies from litter to litter being unimportant in data as extensive as the present group. The variances from which the seasonal correlations were calculated are shown in Table 5. The seasonal correlations (Table 6) were averaged for use in the index by Fisher's "z" transformation for intraclass correlations as follows:

\[ z_i = \frac{1}{2} \log_e \frac{1 + (k_i - 1)r_i}{1 - r_i}, \]

\[ w_k = \frac{2(k_i - 1)(n_i - 1)}{k_i}, \]

\[ \bar{z} = \frac{w_1 z_1 + w_2 z_2 + \cdots + w_6 z_6}{w_1 + w_2 + \cdots + w_6}. \]

The formula for the correlation coefficient between the weight and score of litter mates,

\[ r_{ws} = \frac{\text{cov(between litters)} - \text{cov(total)}}{(k-1)\Sigma w_s}, \]

where cov represents the mean covariance for the designated source, was derived from the path coefficient formula,

\[ r_{ws} = \frac{r_{ws} + (k-1)r_{ws}}{\sqrt{1 + (k-1)r_{ws}} \sqrt{1 + (k-1)r_{ws}}}, \]
Table 5. Between and Within Litter Variances and Covariances For Weight and Score by Seasons.

<table>
<thead>
<tr>
<th>Season</th>
<th>Source of variance</th>
<th>D/f</th>
<th>Mean square</th>
<th>Covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Weight</td>
<td>Score</td>
</tr>
<tr>
<td>1937 F</td>
<td>Between litters</td>
<td>11</td>
<td>2222.0</td>
<td>66.3</td>
</tr>
<tr>
<td></td>
<td>Within litters</td>
<td>49</td>
<td>379.3</td>
<td>15.9</td>
</tr>
<tr>
<td>1938 S</td>
<td>Between litters</td>
<td>77</td>
<td>2971.8</td>
<td>60.6</td>
</tr>
<tr>
<td></td>
<td>Within litters</td>
<td>313</td>
<td>830.1</td>
<td>26.0</td>
</tr>
<tr>
<td>1938 F</td>
<td>Between litters</td>
<td>19</td>
<td>2591.3</td>
<td>52.9</td>
</tr>
<tr>
<td></td>
<td>Within litters</td>
<td>77</td>
<td>987.1</td>
<td>23.0</td>
</tr>
<tr>
<td>1939 S</td>
<td>Between litters</td>
<td>82</td>
<td>2459.0</td>
<td>35.0</td>
</tr>
<tr>
<td></td>
<td>Within litters</td>
<td>360</td>
<td>708.2</td>
<td>19.4</td>
</tr>
<tr>
<td>1939 F</td>
<td>Between litters</td>
<td>21</td>
<td>1059.7</td>
<td>20.6</td>
</tr>
<tr>
<td></td>
<td>Within litters</td>
<td>130</td>
<td>622.3</td>
<td>16.7</td>
</tr>
<tr>
<td>1940 S</td>
<td>Between litters</td>
<td>72</td>
<td>1744.0</td>
<td>26.7</td>
</tr>
<tr>
<td></td>
<td>Within litters</td>
<td>302</td>
<td>639.3</td>
<td>10.0</td>
</tr>
</tbody>
</table>
where $r_{ws}$ is the correlation between the average weight and average score of the same litter. The seasonal correlations were calculated (table 6) and averaged in the same manner as the intraclass correlations above.

The seasonal correlations between productivity of the dam and weight or score of the offspring (shown in table 6) were calculated by weighting the individual productivities according to the number of pigs per litter. The formula used was

$$r_{wp} = \frac{\sum k_i (w_i - \bar{w})(p_i - \bar{p})}{\sqrt{\sum k_i (p_i - \bar{p})^2 \sum k_i (w_i - \bar{w})^2}}.$$  

A summary of the analysis of variance is not given because the correlations were generally small and lacking in significance. The degrees of freedom for each seasonal correlation depended upon the number of litters for that season rather than upon the number of pigs. Although neither the seasonal correlations nor their average values were statistically significant, the average values constitute the most probable values of the parameters which they were intended to estimate and hence were used in the construction of the index.

The phenotypic correlations shown in table 6 were only part of those needed for the determinant of equation (6). Those involving the average weight or score for the litter were derived by means of path coefficients. Since the number
Table 6. The Phenotypic Correlations by Seasons and Their Average Values for Use in the Index.

<table>
<thead>
<tr>
<th>Season</th>
<th>$r_{ws}$</th>
<th>$r_{wp}$</th>
<th>$r_{sp}$</th>
<th>$r_{ww}$</th>
<th>$r_{ss}$</th>
<th>$r_{ws}'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1937 F</td>
<td>.5874</td>
<td>-.0276</td>
<td>-.1573</td>
<td>.4884</td>
<td>.3883</td>
<td>.3870</td>
</tr>
<tr>
<td>1938 S</td>
<td>.6995</td>
<td>.0126</td>
<td>.0444</td>
<td>.3399</td>
<td>.2102</td>
<td>.2013</td>
</tr>
<tr>
<td>1938 F</td>
<td>.6942</td>
<td>-.1603</td>
<td>.0184</td>
<td>.2510</td>
<td>.2120</td>
<td>.0449</td>
</tr>
<tr>
<td>1939 S</td>
<td>.5714</td>
<td>-.0631</td>
<td>-.2356</td>
<td>.3396</td>
<td>.1592</td>
<td>.1365</td>
</tr>
<tr>
<td>1939 F</td>
<td>.4778</td>
<td>-.0301</td>
<td>.0645</td>
<td>.2144</td>
<td>.0406</td>
<td>.2882</td>
</tr>
<tr>
<td>1940 S</td>
<td>.5959</td>
<td>.0188</td>
<td>.0068</td>
<td>.2714</td>
<td>.2664</td>
<td>.1992</td>
</tr>
<tr>
<td>Average</td>
<td>.6143</td>
<td>-.0236</td>
<td>-.0515</td>
<td>.3135</td>
<td>.2043</td>
<td>.1858</td>
</tr>
<tr>
<td>Standard error of average</td>
<td>.0160</td>
<td>.0592</td>
<td>.0591</td>
<td>.0278</td>
<td>.0261</td>
<td>.0256</td>
</tr>
</tbody>
</table>
of pigs per litter \((k)\) is variable, their magnitude depends upon the number of observations per litter:

\[
r_{ww} = \frac{1 + (k-1)r_{ww}}{k} = \frac{1 + (k-1)\cdot 3.135}{k}
\]

\[
r_{ss} = \frac{1 + (k-1)r_{ss}}{k} = \frac{1 + (k-1)\cdot 2.043}{k}
\]

\[
r_{ws} = \frac{r_{ws} + (k-1)r_{ws}^*}{k[1 + (k-1)r_{ss}]} = \frac{0.6143 + (k-1)\cdot 1.858}{k[1 + (k-1)\cdot 2.043]}
\]

\[
r_{sw} = \frac{r_{ws} + (k-1)r_{sw}^*}{k[1 + (k-1)r_{ww}]} = \frac{0.6143 + (k-1)\cdot 1.858}{k[1 + (k-1)\cdot 3.135]}
\]

\[
r_{pw} = \frac{r_{pw}}{1 + (k-1)r_{ww}} = -0.0236 \sqrt{\frac{k}{1 + (k-1)\cdot 3.135}}
\]

\[
r_{ps} = \frac{r_{ps}}{1 + (k-1)r_{ss}} = -0.0815 \sqrt{\frac{k}{1 + (k-1)\cdot 2.043}}
\]

\[
r_{ws} = \frac{r_{ws} + (k-1)r_{ws}^*}{\sqrt{[1+(k-1)r_{ww}][1+(k-1)r_{ss}]}} = \frac{0.6143 + (k-1)\cdot 1.858}{\sqrt{[1+(k-1)\cdot 3.135][1+(k-1)\cdot 2.043]}}
\]

b. Genotypic constants. Of the several methods used for estimating the heritability of characteristics (Lush, 1941), the intrasire regression of offspring on dam was best suited to the present group of data because selection had been practiced among the dams and several of the
lines were one-sire lines. A slight loss of information resulted from the use of this method because a few sires were represented by only one litter. The estimate of heritability.

Table 7. Analysis of Variance and Covariance and the Regression Coefficients for Weight of Offspring on Weight of Dams.

<table>
<thead>
<tr>
<th>Season</th>
<th>D/f</th>
<th>Mean square variance</th>
<th>Covariance</th>
<th>( b_{w_2w_1} )</th>
<th>( w_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Dams</td>
<td>Offspring</td>
<td>Dam/offspring</td>
<td></td>
</tr>
<tr>
<td>1937 F</td>
<td>9</td>
<td>67</td>
<td>1530</td>
<td>29.0</td>
<td>.3328</td>
</tr>
<tr>
<td>1938 S</td>
<td>68</td>
<td>317</td>
<td>2840</td>
<td>60.3</td>
<td>.1904</td>
</tr>
<tr>
<td>1938 F</td>
<td>14</td>
<td>127</td>
<td>1916</td>
<td>75.4</td>
<td>.5932</td>
</tr>
<tr>
<td>1939 S</td>
<td>67</td>
<td>327</td>
<td>2291</td>
<td>109.8</td>
<td>.3361</td>
</tr>
<tr>
<td>1939 F</td>
<td>19</td>
<td>180</td>
<td>929</td>
<td>-57.3</td>
<td>-.3185</td>
</tr>
<tr>
<td>1940 S</td>
<td>55</td>
<td>348</td>
<td>1612</td>
<td>-2.5</td>
<td>-.0072</td>
</tr>
</tbody>
</table>

All seasons \[.1492\]

obtained by doubling the intrasire regression of offspring on dam, represents the heritability of the variance within groups of dams mated to the same sire and contains a small part of the epistatic variance. Since only dams within the same line were mated to a sire, the variance between lines could be conveniently eliminated with that between sires.

The intrasire variance for weight of dams and the covariance for dam and offspring for the 6 seasons are shown
in table 7. Because the dams were much more highly selected for fall litters than for spring litters, there is considerable difference in the variance for dams in the different seasons. The seasonal regression coefficients of offspring on dam differ considerably from season to season, but those showing the greater variability are based on the smaller groups of data. Since dams were sometimes used in successive seasons and offspring of one season were dams in later seasons, there is considerable correlation between the sampling errors for different seasons; hence, the usual formula for the standard error of the difference between two regression coefficients gives too large an estimate for testing the significance of the difference between seasonal estimates.

The seasonal regression coefficients were weighted according to the reciprocal of their respective variances,

\[ V(n-1) = \frac{\sum k_i (w_1 - \bar{w}_1)(w_2 - \bar{w}_2)}{\sum k_i (w_1 - \bar{w}_1)^2}, \]

\[ w_1 = \frac{\sum k_i (w_1 - \bar{w}_1)^2}{V}, \]

\[ \sum \frac{w_2 w_1}{w_1 + w_2 + \cdots + w_6} = .1492, \]

where \( w_1 \) and \( w_2 \) represent the weight of the dam and the average weight of the pigs in the litter, respectively, \( k \) the number of pigs in the litter and \( n \) the number of litters. The average regression coefficient for the weight of offspring on that of
dams was slightly smaller than most of those found by What-
ley (1939) who used several different methods of estimation.
Whatley's estimates were calculated from data on the Iowa
Staton swine herd from the spring of 1934 through the spring
of 1938, so that the data from 452 of the pigs used in this
study also were used in his study. This estimate must be
doubled to give the heritability of weight at 180 days of
age, the value .2954 being used in the construction of the
index.

Because the scoring system was not introduced until
the spring of 1937, a number of litters could not be used in
the estimation of the heritability of score, as the scores
of their dams were not available. All of the data taken on
the pigs born in the fall of 1937 and some of that for pigs
born in 1938 had to be omitted for this reason. Otherwise
the heritability of score was calculated in a manner identi-
cal with that of the heritability of weight. The results are
shown in table 8. The regression coefficient doubled gives
.0978 for the estimate of the heritability of score.

Since breeding stock was saved only from spring
litters, with the exception of a few animals in the fall of
1939, and only daughter-dam pairs, both of which had produced
litters, could be used in estimating the heritability of pro-
ductivity, the data were very limited. The same method of
calculating the seasonal regression coefficients and weight-
Table 8. Analysis of Variance and Covariance and the Regression Coefficients for Score of Offspring on Score of Dams.

<table>
<thead>
<tr>
<th>Season</th>
<th>D/f</th>
<th>Mean square variance</th>
<th>Covariance</th>
<th>( b_{2.5} )</th>
<th>( W_{1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Dams</td>
<td>Offspring</td>
<td>Dam/offspring</td>
<td></td>
</tr>
<tr>
<td>1938 S</td>
<td>42</td>
<td>17.1</td>
<td>45.0</td>
<td>-1.51</td>
<td>-.0882</td>
</tr>
<tr>
<td>1938 F</td>
<td>6</td>
<td>13.9</td>
<td>17.5</td>
<td>2.06</td>
<td>.1482</td>
</tr>
<tr>
<td>1939 S</td>
<td>67</td>
<td>10.9</td>
<td>34.8</td>
<td>1.27</td>
<td>.1168</td>
</tr>
<tr>
<td>1939 F</td>
<td>19</td>
<td>3.1</td>
<td>19.1</td>
<td>0.64</td>
<td>.2065</td>
</tr>
<tr>
<td>1940 S</td>
<td>55</td>
<td>5.2</td>
<td>24.4</td>
<td>-0.05</td>
<td>-.0101</td>
</tr>
</tbody>
</table>

All seasons


<table>
<thead>
<tr>
<th>Season</th>
<th>D/f</th>
<th>Mean square variance</th>
<th>Covariance</th>
<th>( b_{2.5} )</th>
<th>( W_{1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Dams</td>
<td>Daughters</td>
<td>Dam/offspring</td>
<td></td>
</tr>
<tr>
<td>1938 S</td>
<td>28</td>
<td>48.5</td>
<td>181</td>
<td>5.98</td>
<td>.1233</td>
</tr>
<tr>
<td>1939 S</td>
<td>29</td>
<td>58.2</td>
<td>93</td>
<td>3.78</td>
<td>.0650</td>
</tr>
<tr>
<td>1940 S</td>
<td>25</td>
<td>50.1</td>
<td>99</td>
<td>3.94</td>
<td>.0786</td>
</tr>
</tbody>
</table>

All seasons

.0806
ing them according to the reciprocal of their respective variances to obtain an average estimate for all seasons was followed. The data are summarized in table 9. The heritability of productivity was estimated as .1612 from the doubled regression coefficient.

Rather surprisingly the heritability of productivity was the least variable from season to season of any of the three characters, although this figure was the least reliable statistically. Standard errors were not calculated for these estimates because the seasonal values were not independent of each other, the figures for some animals appearing for dams in more than one season or for offspring in one season and for dams later. If the seasonal estimates had been independent, the average estimate in each case would have the standard error

\[ S.E.(b) = \frac{1}{\sqrt{\frac{1}{W_1} + \frac{1}{W_2} + \ldots}} \]

where the \( W \)'s represent the reciprocals of the seasonal variances as shown in the preceding tables. Of the three, it is certain that the heritability of score and productivity are not based on enough data to be statistically significant. Hence, future investigations may prove that these estimates are rather far from the parameters they are intended to estimate.*

---

See page 98 for a discussion of the evidence from previous investigations regarding the heritability of these characters.
If they are all proportionately too low or too high, no error will result in the relative amounts of attention to be paid the various characteristics.

The role of genotypic as compared to environmental correlations in causing observed phenotypic correlations between the characteristics of an animal was given particular attention in a previous section. An investigation of the biological basis for the two types of correlations indicated that the environment peculiar to an animal (and random with respect to any one characteristic) was a potential source of correlation between characters. No statistical technique is available for the direct analysis of an observed phenotypic correlation into its residual sources. Genotypic correlations can, however, be derived by resorting to the correlations between relatives which have been randomized with respect to environment, just as estimates of heritability are derived in this manner.

The accuracy of a genotypic correlation derived from the correlation between relatives depends in part upon the genetic relationship of the relatives; hence, it is desirable for purposes of estimation to have as high degree of relationship between the relatives as possible. In the present group of data those contemporary relatives most closely related were litter mates and were therefore exposed to similar environmental deviations, being developed in the same
uterus, suckled by the same sow, reared to weaning in the same lot, etc. Parent-offspring correlations were unsuitable for measuring genotypic correlations involving weight or score in the present group of data, since the parents represented a selected group among their generations.

Regressions of offspring on parent when the parents are selected afford an unbiased estimate of that regression for which the selected characteristic is the independent variable. Since an unbiased estimate of the correlation is the geometric mean of the two unbiased regression coefficients, which can be calculated, the genotypic correlation can be derived from regressions of offspring on dam. For any pair of characters two unbiased regressions of offspring on dam exist: (1) The regression of character A of the offspring on character B of the dams and (2) the regression of character B of the offspring on character A of the dams. Therefore,

$$\sqrt{r_{A_2B_1}\cdot r_{B_2A_1}} = r_{A_1B_2} = r_{B_1A_2}$$

where the subscripts 1 and 2 represent the dams and their offspring, respectively. By this method, estimates of the correlation such as would exist in populations undisturbed by selection can be obtained. Moreover, all of the information is utilized in the estimate.

If parents and offspring are randomized with respect to environment, a correlation between one character of the parents and the other character of the offspring must be genetic
in origin. That is,

\[ r_{A_1 B_2} = r_{B_1 A_2} = \frac{1}{2} \epsilon_{A B} G_A G_B. \]

The consequences of dominance or factor interaction with respect to genotypic correlations have not been worked out in detail, but it appears certain that a genotypic correlation calculated by the above method contains a small fraction based on the covariance of epistatic genes which influence both characters in some combinations but not in others. The question of mating systems and genetic differences between dams mated to different sires can be avoided by leaving their effects unanalyzed; i.e., by the use of intrasire regressions.

The sampling error of a genotypic correlation calculated in this manner is of necessity very large. In the first place, four statistics must be calculated from the data. Secondly, none of these statistics are entirely independent of any of the others so some correlation exists between their sampling errors. However, the method appears to utilize all the information available and to give an unbiased estimate, thus qualifying as the best statistic for the purpose.

The regression coefficients for weight of offspring on score of dam and for score of offspring on weight of dam are shown by seasons in table 10. The averages for each were obtained by weighting the seasonal estimates by the reciprocal of their respective variances \( W_1 \). The seasonal regressions
are prevalingly positive, and the three exceptions are thought to be due to sampling errors. The statistics based on the data taken in the fall of 1939 have large sampling errors because

Table 10. Seasonal Regression Coefficients for Score of Offspring on the Weight of Dams and for Weight of Offspring on the Score of Dams.

<table>
<thead>
<tr>
<th>Season</th>
<th>$b_{w1s2}$</th>
<th>$W_1$</th>
<th>$b_{w2s1}$</th>
<th>$W_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1937 F</td>
<td>0.0045</td>
<td>101</td>
<td>0.0437</td>
<td>1.42</td>
</tr>
<tr>
<td>1938 S</td>
<td>0.0072</td>
<td>2564</td>
<td>0.04794</td>
<td>0.34</td>
</tr>
<tr>
<td>1938 F</td>
<td>0.0904</td>
<td>961</td>
<td>0.8753</td>
<td>1.91</td>
</tr>
<tr>
<td>1939 S</td>
<td>0.0286</td>
<td>3668</td>
<td>0.8753</td>
<td>1.91</td>
</tr>
<tr>
<td>1939 F</td>
<td>-0.0278</td>
<td>1334</td>
<td>-2.0852</td>
<td>0.56</td>
</tr>
<tr>
<td>1940 S</td>
<td>-0.0021</td>
<td>4524</td>
<td>0.7107</td>
<td>1.04</td>
</tr>
<tr>
<td>All seasons</td>
<td>0.0128</td>
<td>1.534</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

the fall pig crops were small, and the negative regression found for the spring of 1940 in one case is practically zero. However, neither of the average regression coefficients is significant so the true values could be zero or even negative.

The correlation expected in unselected populations is thus

$$r_{w1s2} = r_{s1w2} = \sqrt{(0.0128)(1.534)} = 0.044,$$
whereas, when calculated directly, \( r_{w_1s_2} = .027 \) and \( r_{s_1w_2} = .041 \). These results are in agreement with the expectation that selection would result in a smaller correlation than that observed in unselected populations.

The statistics thus far calculated afford a direct estimate of the genotypic correlation between score and weight,

\[
r_{G,G_S} = \frac{b_{w_1s_2} - b_{w_1s_1}b_{s_1w_2}}{b_{w_1s_2}b_{s_1s_2}} = .5186 .
\]

This relatively large genotypic correlation is made plausible by the fact that factors (either genetic or environmental) which affect the vigor, health and general constitution of the animal are thought to act physiologically in the same direction on score and weight, although weight is measured at a constant age and score is taken at a constant weight.

The correlation between productivity of the dam and the 180-day weight of the pigs in the litter upon which the estimate of productivity was based was -.0204. Since it was uncertain whether or not the above correlation contained some environmental effect, the correlation of .0241 between the weight of the pigs in one litter and the productivity of the dam based on another litter was calculated. Theoretically, the first is unbiased as far as the effects of selection are concerned but might contain some environmental correlation, whereas the second might be too small because of selection.
for productivities, since only dams having two or more litters were used in this correlation. The correlation of .0079 between the weight of the dam and productivity of the offspring might be biased in the direction of smallness by selection for heavy dams. The sampling errors are much too large to attach statistical significance to any of these correlations.

The genotypic correlations between weight and productivity derived from the three correlations, in the order given above, are -.1860, .2198 and .072, respectively. Although it may be maintained that the bulk of the evidence indicates a positive correlation, this is by no means certain, and, since neither the direction nor magnitude (other than that it is small) were clearly defined, it was thought best to assume that the real genotypic correlation between weight and productivity was zero.

The corresponding phenotypic correlations for productivity and score were -.0733, -.0220 and .1254, respectively, none of which are statistically significant. Estimates of the genotypic correlations derived from them are -1.1672, -.3822 and 1.9968, the first and last of which are impossibly large. The situation here is worse than that above, as it cannot even be said with certainty that the true genotypic correlation is likely to be small.

Because the average of the estimates obtained was near zero and because it offered the possibility of least
error, the value of zero was assumed for the genotypic correlation between score and productivity.

The genotypic constants required in equation (9) having been calculated from the data, the correlations between components of the index were calculated as follows:

\[ r_{HW} = \frac{d_{G_w} + d_{G_s}}{G_{w_s}} = .3495, \]

\[ r_{HS} = \frac{d_{G_w} G_{w_s} + d_{G_s}}{G_{w_s}} = .1370, \]

\[ r_{Hp} = \frac{d_{G_p} M_p}{M_p} = .3046, \]

\[ r_{Hs} = \frac{r_{HS} \left( 1 + (k-1)q \right)}{k \left( 1 + (k-1)r_{ss} \right)} = .1370 \frac{1 + (k-1)q}{k \left( 1 + (k-1).2043 \right)}, \]

\[ r_{HW} = \frac{r_{HW} \left( 1 + (k-1)q \right)}{k \left( 1 + (k-1)r_{ww} \right)} = .3495 \frac{1 + (k-1)q}{k \left( 1 + (k-1).3135 \right)}, \]

where \( q \) and \( M \) represent the coefficient of relationship for litter mates and for dam and offspring, respectively.

5. The regression coefficients

After the cofactors (C's) of the determinant (equation 6) and the correlations between each of the five components of the index and the aggregate genotypic value (equation 9) were calculated, the standard regression coefficients were derived by use of equation (5). Only two of them are simple enough to be shown in general form. These are:
\[
B_w = \frac{(1-q)[r_{HW}(1-r_{sS}) - r_{HS}(r_{ws}-r_{ws}')]^2}{(1-r_{ww})(1-r_{sS}) - (r_{ws}-r_{ws}')^2},
\]
\[
B_s = \frac{(1-q)[r_{HS}(1-r_{ww}) - r_{HW}(r_{ws}-r_{ws}')]^2}{(1-r_{ww})(1-r_{sS}) - (r_{ws}-r_{ws}')^2},
\]

where \(q\) is the coefficient of relationship between litter mates.

The phenotypic and genotypic constants were substituted in the general formulae and the following equations obtained:

\[B_w = (1-q)(.6049),\]
\[B_s = -(1-q)(.1535),\]

\[B_w = \sqrt{k[1+(k-1),5035] \cdot [0.0915(k-1)-.4269+q \cdot [3.939+.2270(k-1)]]} \cdot \frac{2.3038+1.0562(k-1)+.1047(k-1)^2}{2.3038+1.0562(k-1)+.1047(k-1)^2},\]

\[B_s = \sqrt{k[1+(k-1),2043] \cdot [1.058+.0281(k-1)-q \cdot [3.536+.0958(k-1)]]} \cdot \frac{2.3038+1.0562(k-1)+.1047(k-1)^2}{2.3038+1.0562(k-1)+.1047(k-1)^2},\]

\[B_p = \frac{.3578+.1639(k-1)+.0170(k-1)^2-q(k-1) \cdot [0.003+.0026(k-1)]}{2.3038+1.0562(k-1)+.1047(k-1)^2}.\]

The "betas" were multiplied by the factor, \(3.074\frac{G_H}{G_{x_1}}\), in converting them to regression coefficients, in order to increase the standard deviation of the index to a more convenient size for keeping records. The effect of the factor, 3.074, is very nearly the same as would have resulted if the
relative values (the a's) of the three characteristics had been referred to 180-day weight as a basis instead of to score. The chief advantage of this exact factor is that, by using it, three of the regression coefficients become even fractions.

The coefficient of relationship between litter mates (q) has been left in general form in the formulae to this stage because it is determined by the mating system. Wright (1921b) showed that (1) in parmiotic populations q = .5; (2) under assortive mating q may become slightly larger than .5; (3) under disassortive mating q may become slightly less than .5; (4) in populations consisting of a number of inbred lines q is greater than .5, depending upon the inbreeding coefficient. Lush (1940b) found that in populations consisting of single inbred lines q depended upon the number of males and females which became parents in the line. For example, in a one-sire line with an unlimited number of females q = .275 on an intraline basis. For more than two sires and a large number of females, q is nearly .5, approaching that figure asymptotically as the number of sires increases. On an intraline basis, the relationship of full sibs remains at .5 regardless of the level of inbreeding until perfect homozygosity is reached.

Although some of the lines from which the data of this study were taken were intended to be one-sire lines,
most of these have generally had two or more sires represented in each crop of pigs (due to the ever-present risk of losing sires during the breeding season and the temporary necessity of maintaining positive and negative branches in combatting disease). Hence, the actual regression coefficients shown below were calculated for $q = .5$. The regression coefficients are

$$b_w = .3$$
$$b_s = -.5$$
$$b_p = .5$$

$b_w$ and $b_s$ are given for different values of $k$ in table 11.**

Since negative values for the index are inconvenient in keeping records and tentative indexes are sometimes desirable before the average weight and score are available, the index given below was chosen for its accuracy and utility:

$$I = .3\tilde{W} - .5\tilde{S} + .5P + b_w(\tilde{W} - \bar{W}) - b_s(\tilde{S} - \bar{S}),$$

where $\tilde{W}$ and $\tilde{S}$ represent the average weight and score for all the pigs in a line that season. The average of this index will usually be between 40 and 50, thus requiring a pig to be three or four standard deviations below the average to have a negative index. Tentative indexes can be calculated, before

**The size of $k$ had so little effect upon $b_p$ that the error introduced by assuming it constant is unimportant.
Table 11. Values of $b_1$ and $b_3$ for Values of $k$
From 1 to 13.

<table>
<thead>
<tr>
<th>$k$</th>
<th>$b_1$</th>
<th>$b_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.116</td>
<td>-.204</td>
</tr>
<tr>
<td>2</td>
<td>.180</td>
<td>-.346</td>
</tr>
<tr>
<td>3</td>
<td>.221</td>
<td>-.453</td>
</tr>
<tr>
<td>4</td>
<td>.249</td>
<td>-.537</td>
</tr>
<tr>
<td>5</td>
<td>.270</td>
<td>-.605</td>
</tr>
<tr>
<td>6</td>
<td>.287</td>
<td>-.661</td>
</tr>
<tr>
<td>7</td>
<td>.300</td>
<td>-.708</td>
</tr>
<tr>
<td>8</td>
<td>.311</td>
<td>-.743</td>
</tr>
<tr>
<td>9</td>
<td>.320</td>
<td>-.782</td>
</tr>
<tr>
<td>10</td>
<td>.327</td>
<td>-.812</td>
</tr>
<tr>
<td>11</td>
<td>.333</td>
<td>-.838</td>
</tr>
<tr>
<td>12</td>
<td>.339</td>
<td>-.862</td>
</tr>
<tr>
<td>13</td>
<td>.344</td>
<td>-.883</td>
</tr>
</tbody>
</table>
the average of a line is known exactly, by using the long-
time average of the line.

6. Test of the index

The important theoretical premises required in the
derivation of the index are (1) that the genotypic and en-
vironmental factors combine additively to produce the observed
phenotypes, (2) that they are independent and (3) that the
quantitative genotypes and phenotypes are normally distrib-
buted. Assumptions comparable to these are so commonly made
in many phases of statistical work that they are conventional.
Only the latter part of the third assumption can be tested
by direct examination of the distributions of the phenotypes.
The others can be examined indirectly by the erection of hy-
potheses to be accepted or rejected depending upon the valid-
ity of the assumptions mentioned above.

As remarked previously, the index was constructed
from the data taken on pigs born from the fall of 1937 through
the spring of 1940, whereas the data from the pigs born in the
fall of 1940 were reserved for a test of the index. The lat-
ter data consisted of records taken on 230 pigs from 49 lit-
ters and 16 sires. These data cannot be considered numerous
in view of the variability associated with the genotypic con-
stants but are sufficient to draw conclusions as to the gen-
eral validity of the index and to indicate paths of further
investigation if those prove necessary. That is, lack of agreement between an expected result derived from the theoretical properties of the index and the result actually observed in the 1940 fall crop of pigs might conceivably be due either to errors inherent in the index or to sampling errors associated with the number of pigs upon which the test must be made.

A comparison of the theoretical and actual results are shown in table 12. The expected $R^2_{\text{IH}}$ was calculated from equation (7) by substituting in the designated constants taken from the 1937-1940 data. The actual $R^2_{\text{IH}}$ was calculated by figuring indexes for the 1940 fall pigs and their dams and computing the regression of the index of the offspring on that of the dams. The expected $\sigma^2_1$ was calculated from the relation $\sigma^2_1 = \sum_{i=1} b_i x_i^2 + 2 \sum_{i<j} r_{ij} x_i x_j$, whereas the actual standard deviation was computed from the indexes of the 1940 fall pigs. The average gains in weight, score and dam's productivity, assuming $p = .5$, for the pigs having the highest indexes were calculated theoretically by the relation $z/p \, x_i^2 / \sigma x_i^2$, and actually by marking the highest half of the indexes in the fall of 1940 and then calculating their average for the three characteristics. The only constant taken from the 1940 fall pigs which was used in the derivation of the theoretical results was $k = 4.694$. 


The first four of the five comparisons between the theoretical and actual results are surprisingly close, only the fifth being seriously off. The explanation for the discrepancy in the actual and expected gain in dam's productivity lies in part in the unexpected uniformity which this characteristic had in the fall of 1940. In the 1940 fall pigs weight and score were more variable than the averages for the 1937 fall-1940 spring seasons, while dam's productivity was less variable. These differences caused weight and score to have more influence upon the index for the 1940 fall pigs than was expected according to the earlier data, so that the pigs having the highest indexes were heavier and scored higher than expected, whereas they were lower in dam's productivity.

Table 12. Actual (1940 Fall) and Theoretical Phenotypic Gains Due to Selecting the Highest Half of the Indexes.

<table>
<thead>
<tr>
<th>Phenotypic gain</th>
<th>Actual</th>
<th>Theoretical</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S_I )</td>
<td>12.92</td>
<td>12.75</td>
</tr>
<tr>
<td>( T_S - I )</td>
<td>10.92</td>
<td>10.20</td>
</tr>
<tr>
<td>( W_S - W )</td>
<td>25.94</td>
<td>21.74</td>
</tr>
<tr>
<td>( S_S - S )</td>
<td>1.5</td>
<td>1.30</td>
</tr>
<tr>
<td>( F_S - F )</td>
<td>.9</td>
<td>2.94</td>
</tr>
<tr>
<td>( R^2_{IH} )</td>
<td>.152</td>
<td>.163</td>
</tr>
</tbody>
</table>
The index, since it is based on several characters, tends toward the stability observed in averages, a fact expected on theoretical grounds. For example, score was more heritable and weight was less heritable in the 1940 fall data than the averages from the earlier data, so that the average as represented by $R^2_{IH}$ was very close to the expected. Also, although the variances for the characters in the 1940 fall data differed somewhat from those found in the earlier data, the standard deviation of the index was near the expected.

7. The comparison of indexes

The most logical basis for the comparison of indexes is offered by a contrast of the expected genetic progress resulting from their use. The aggregate genetic gain expected due to selection was shown to be

$$E(\mathbf{H}) = Z/p \, R_{IH} \, \sigma_H^2$$

where $I$ is the variable which is selected directly and $H$, the aggregate genotypic value, is completely defined as a linear function of the genotypes. Since both $Z/p$ and $\sigma_H$ are independent of the index used, the comparison of indexes can most directly be achieved by contrasting the correlations between each index and the aggregate genotypic value.

Since an index is a linear function of measurable phenotypes and $H$ is a function of the economically valuable genotypes, $R_{IH}$ can be calculated if the relations between the
genotypes and phenotypes are known. That is,

\[ R_{IH} = \left\{ 2a_i b_i \sigma_{x_i} \sigma_{g_i} x_i g_i + 2a_i b_j \sigma_{x_j} \sigma_{g_j} x_j g_j \right\} \frac{1}{\sigma_{F_H}^2} \]

for any particular index in question. In the comparisons which follow it is necessary to assume that the relative values of a unit change in each characteristic (the \( a_i \)'s) are accurately assigned and that the phenotypic and genotypic relations of the system which were calculated from the data are correct. There is no other basis for comparison (other than an arbitrary one). Since small errors certainly exist and it may be presumptuous to trust that all are small, any conclusions drawn from the comparison of indexes are somewhat limited. That is, slightly different indexes give results so nearly similar that a slight advantage in one index might be reversed if some of the constants were actually in error.

Since some of the indexes depend upon the values of \( a_i \) and \( b_i \), a value of \( .5 \) (true in panmictic populations) was assumed for them throughout the comparisons. For those indexes which involve the litter averages for weight and score, a value of \( 5 \) was assumed for \( k \) since the average number of pigs reared does not in general deviate much from this figure.

The indexes to be compared are:

1. The index developed in this study, the coefficients of which make \( R_{IH} \) a maximum.
\[ I_1 = .3W - .5S + .5P + .270W - .605S. \]

(2) An index based on the individual's own weight and dam's productivity and the average weight of the litter, the coefficients of which make \( R_{I_2H} \) a maximum,

\[ I_2 = .246W + .213W + .519P. \]

(3) An index based on the animal's own weight, score and dam's productivity, the coefficients of which make \( R_{I_3H} \) a maximum,

\[ I_3 = .419W - .911S + .492P. \]

(4) An index upon which selections were based in previous years, the regression coefficients of which were assigned according to the economic value of the characteristics,

\[ I_4 = 1/3 W + S + P + .303W + 1.667S. \]

(5) An index based only on the animal's own 180-day weight,

\[ I_5 = W. \]

(6) An index based only on the animal's own score,

\[ I_6 = S. \]

Direct comparison of the indexes as they are given is almost meaningless. The single most informative statistic for purposes of comparison is the correlation between each of the indexes and the aggregate genotypic value as shown in table 13. The changes caused in the average for each characteristic for a selection differential of .5 is also of interest, because these average changes represent the visible effects of selection upon a selected population. The sum of the visible
Table 13. The Changes in Genotypic and Phenotypic Means Incurred From Index Selection as Affected by Different Indexes (Z/p = .5).

<table>
<thead>
<tr>
<th>Expected gain</th>
<th>Index</th>
<th>( I_1 )</th>
<th>( I_2 )</th>
<th>( I_3 )</th>
<th>( I_4 )</th>
<th>( I_5 )</th>
<th>( I_6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenotypic</td>
<td>( E(W) )</td>
<td>21.8</td>
<td>22.3</td>
<td>22.2</td>
<td>21.6</td>
<td>25.5</td>
<td>15.7</td>
</tr>
<tr>
<td></td>
<td>( E(\overline{3}) )</td>
<td>1.35</td>
<td>1.94</td>
<td>1.09</td>
<td>2.62</td>
<td>2.35</td>
<td>3.83</td>
</tr>
<tr>
<td></td>
<td>( E(\overline{F}) )</td>
<td>2.96</td>
<td>2.96</td>
<td>3.08</td>
<td>2.72</td>
<td>-0.18</td>
<td>-0.63</td>
</tr>
<tr>
<td></td>
<td>( E(\overline{K}) )</td>
<td>14.5</td>
<td>15.3</td>
<td>14.7</td>
<td>15.3</td>
<td>10.5</td>
<td>7.8</td>
</tr>
<tr>
<td>Genotypic</td>
<td>( E(G_w) )</td>
<td>7.76</td>
<td>7.15</td>
<td>7.10</td>
<td>6.25</td>
<td>7.61</td>
<td>2.27</td>
</tr>
<tr>
<td></td>
<td>( E(G_s) )</td>
<td>0.27</td>
<td>0.32</td>
<td>0.23</td>
<td>0.36</td>
<td>0.33</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>( E(G_p) )</td>
<td>0.25</td>
<td>0.25</td>
<td>0.24</td>
<td>0.27</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>( E(H) )</td>
<td>10.22</td>
<td>9.87</td>
<td>9.72</td>
<td>9.20</td>
<td>8.80</td>
<td>3.47</td>
</tr>
<tr>
<td>( R_{IH} )</td>
<td></td>
<td>0.404</td>
<td>0.390</td>
<td>0.385</td>
<td>0.364</td>
<td>0.350</td>
<td>0.137</td>
</tr>
</tbody>
</table>
changes in the characteristics, weighted according to the economic value of each character, may be considered as the apparent progress. That is,

\[ E(\overline{X}) = \frac{1}{3} E(W) + E(\overline{S}) + 2E(F) \]

represents the phenotypic gain in the selected group.

The average changes in the genotypes for each character and in the aggregate genotype are of particular interest because these values represent the permanent improvement which the selected group will transmit to their offspring. The latter comparison indicates which characters are affected and the extent to which each is affected permanently by the use of each index. Since the sum of the gains in the three characters, weighted according to their relative economic importance, is equivalent to the gain in the aggregate genotypic value,

\[ E(H) = \left\{ \frac{1}{3} E(G_w) + E(G_s) + 2E(G_p) \right\} \times 3.074, \]

the constant 3.074 appearing only because in the calculations each index was referred to the same \( H \) as that of \( \Delta_1 \).
IV. DISCUSSION

The data were sufficiently numerous to provide fairly accurate estimates of the population parameters for the phenotypic correlations shown in table 6. Although indications that the phenotypic variances for score and weight differ from one season to another were definite, the correlations between the characters appeared much more stable. To presume that the constants found in the Iowa Station herd are representative of those in herds under different conditions and mating systems is probably unsafe, especially for score, without actually comparing them with results from studies of other populations. Score is a highly subjective measurement, differing from weight and productivity in that the attitudes and ideals of the men who act as judges are largely responsible for the variability and form of the distribution of scores. The scoring system used in the Regional Swine Breeding Laboratory was adopted only after studies (Lush and Craft, 1937, 1938; Lush, 1938c) showed that scorers who had approximately the same kind of technical training in livestock judging agreed with each other fairly well and held similar ideals. The collaborators at different stations endeavor to adhere to the same ideals by communication with each other and exchange of ideas. Attempts to replace the scoring system by schemes of exact body measurement have been unsuc-
cessful as a means of describing market desirability; hence, a scoring or judging system of some kind will probably remain as one of the characteristics of any selection program for livestock.

The data were less satisfactory in regard to the reliability of the estimates for the genotypic parameters. The variances of these estimates were given in order that later estimates could be added to those obtained herein. Rather wide fluctuations were noted in the seasonal estimates for the heritability of score and weight, whereas the seasonal estimates for heritability of productivity, which were least accurate statistically, were reasonably constant. The figure found for the heritability of weight substantiated that reported by Whatley (1939) who estimated that between 30 and 40 percent of the variance was hereditary. Stonaker (1941) and Craft (1941) reported that about 20 to 30 percent of the variance in score is heritable, estimates which are within the fiducial limits of the 10 percent found in this study. Although the heritability of productivity had not been studied previously, Lush and Molln (1941) found that about 15 percent of the variance in number of pigs farrowed and in weight of the litter at 56 days of age was attributable to permanent differences among sows. If most of the permanent differences are due to the additive effects of genes, the estimate of 16 percent found in this study for the heritability of productivity is not considered an unreasonable figure, since
dealt, it was not examined carefully because the assumptions

the theoretical base of the index was examined in considerable

principles when permit maximum genetic progress. Although

these estimates of on index constructed according to the

estimation is useful. These programs of estimation, however, are

score and productivity is either a factor large positive or

estimated and unless the true genetic correlation between

errors in the statistic used in the construction of the index

From general considerations it appears that serious

sumed only as a course of least errors under the circumstances.

construction of the index was a matter of choice and was as

interpretation. The value of zero assumed to the statistic in the

since neither the magnitude nor the sign was established definite

duetoart can hardly be regarded otherwise than as a failure

the value of the genetic correlation between score and pro-

been either positive or negative. The attempt to estimate

and productivity appeared to be the better since but could have

little value, apparently about .5 which between weight

weight and score was established relatively well. Astatistic

with only partial success. The genetic correlation between

errors to estimate the genetic correlations met

on the weight or the traits at different ages.

productivity is a linear function of the number of猪 and
involved have long since become commonplace in the statistical analysis of genetic populations. The only new features required were (1) the evaluation of unit changes in different characters according to their economic importance and (2) the analysis of phenotypic correlations into their genotypic and environmental sources.

Although it was statistically possible to calculate sampling errors for each of the constants used in the index and thus for the index itself so that a precise statistical test could be devised to examine its accuracy, the data were numerically insufficient to provide both reliable estimates of the constants required and an independent test of the accuracy of the index. The more precise test was sacrificed in order to establish with greater accuracy the constants upon which the index was based. The test devised was that of calculating actual indexes for the pigs of the 1940 fall season, the records of which were not used in the construction of the index, and of comparing the actual effects of selection with those expected for theoretical reasons. While the test cannot be considered a statistical one insofar as the definition of limits for the acceptance or rejection of hypotheses is concerned, it is evident that in most properties tested the theoretical and actual results agreed remarkably well.

The criterion of maximum genetic progress appears to have wide application as a basis to which many problems of
For the same reason by the negative regression coefficient for the first statement and correlated in the actual index of uncorrelated correlation between characters, it is stated later by an hypothesis application was forecast in the study of the genetic basis for a phenotypic correlation estimate between the two traits. Thus, what is demonstrated determines pecuniary to an economic character when made of an unimportant character as an indicator of the en-

The most unimportant result of this study is the use

that is considered

consideration the result only when the non-linear type of use

that for cases where the assumption is not met of possible assumptions are less easily verified, and it may be possible to the accuracy of errors as inferred from these genetic experiments under correlation. In the present era, this might be overcome, and if this were done for a study of an hypothetical situation, genetic progress could be achieved. Although these appear to be no reason for assuming be manifested as a breeder's ideal, the pedigrees, never considered, correlation between the selected and unselected, members of the

the breeding problem can be referred to through these criteria.
score. The "beta" for score (page 85) given in general form
would reduce to

\[ B_a = \frac{Hs - Hw^*ws}{1 - r_{ws}^2}, \]

for an index based on the pig's own score and weight only, a
form which can be studied more easily. This formula shows
that the combination of circumstances which cause an economic
character to have a negative regression coefficient (even
though it has a positive genotypic correlation with other
economic characters) are:

1. The character must be relatively unimportant econom-
ically and/or show relatively little genetic variabi-

ity.

2. There must be a high phenotypic correlation with a
more influential character.

The circumstances causing negative regression coefficients in
the more complicated index are a little more complex since the
correlation between weights and between weight and score of
litter mates enters the picture.

Since the statistical relations between the genotypes
and phenotypes for weight and score of litter mates have been
calculated, the environmental correlations between the charac-
teristics of the same pig and of his litter mates could be de-
duced. These are:
\[
\begin{align*}
\rho_{E_{w_1}E_{s_1}} &= .6608, \\
\rho_{E_{w_1}E_{s_2}} &= .2342, \\
\rho_{E_{w_2}E_{s_2}} &= .1779, \\
\rho_{E_{s_1}E_{s_2}} &= .1722, \\
\end{align*}
\]

where the subscripts 1 and 2 refer to pigs which are litter mates. The correlation of .6608 between the environmental factors which influence score and weight of the same pig is surprisingly large. It should be remembered that this correlation may contain some of the correlation between dominance and epistatic deviations of the pleiotropic genes which affect each of the characters.

Considerable insight as to how indexes attain their efficiency can be obtained by comparing the genetic gains expected for each of the individual characters of the breeding program at a given level of selection. These gains were calculated for six indexes (table 15) to indicate the relative changes in both genotypes and phenotypes caused by direct selection for different indexes.

The first index, \(I_1\) (page 94), can be taken as a basis for comparison since it was the most efficient of the six indexes. \(I_2\) was constructed according to the same principles as \(I_1\), the difference in their efficiency being due to
loss of information because score was eliminated. It is interesting that there was an actual increase in the genetic gain for score for $I_2$, and the actual loss of information came about through a decrease in the genetic gain for weight. This again indicates that the primary function of score in the index is as an indicator of the environment for weight. $I_2$ is 96.5 percent as efficient as $I_1$. This difference can be made a basis for retaining or eliminating score from the breeding program, depending upon the time and effort required in taking and recording scores.

$I_3$ was constructed according to the same principles as the two previous indexes but is based only on the weight and score of each pig and on the productivity of his dam. Rather surprisingly, it is less efficient than $I_2$, although it contains score. Some of the genetic gain possible by use of $I_1$ or $I_2$ in each of the three characters is lost by the use of $I_3$. The efficiency of $I_3$ is 95.3 percent of that of $I_1$.

The comparison of $I_1$ and $I_4$ is especially interesting because they contain the same information. The difference in the two indexes arises from the amount of attention paid each character. The use of $I_4$ results in considerable loss of gain in the genotype for weight, chiefly because the index does not take advantage of the information in score concerning weight. $I_4$ is 90.1 percent as efficient as $I_1$. 
I₅, an index based only on weight, again indicates the value of a correlated character in evaluating the genotype of a second, since the genetic gain in weight is larger for I₁ than for I₅, although part of the gain in I₁ is due to the use of the litter average for weight. Although the dams of the selected pigs are lower than the average of their contemporaries in productivity $E_5(F) = -0.18$, selection causes no genetic loss in productivity. I₅ is 86.6 as efficient as I₁, indicating a difference of only 13.4 percent between the best index possible with the five variables and one based on weight alone. In the present case, weight is the most highly heritable as well as the most variable of the characteristics. It would be very easy to set up an arbitrary criterion of selection based upon the three characters the efficiency of which was less than that of I₅.

I₆ was brought into the comparisons to show that (1) the relative efficiency of an index based on a character popularly considered very important is relatively low and (2) most of the improvement in the aggregate genotype expected from its use is due to genetic gain in weight and not to score. The efficiency of I₆ is only 33.9 percent of that of I₁.

These indexes are not considered to be less efficient than similar indexes which might be constructed for other types of livestock. The ideal situation, of course, would be to have an index whose correlation with the aggregate
genotype was 1.00. However, the more complicated an index becomes the more opportunities are made for errors in some of the constants upon which it is based. There does appear, however, to be real opportunity for increasing the rate of genetic progress by the proper construction of indexes from information which is commonly recorded by careful breeders. For example, the efficiency of the index could be improved several percent by including the average performance of half-sibs for score and weight because (1) there is probably no environmental correlation between half-sibs, and (2) there are usually several litters sired by the same male.
V. SUMMARY AND CONCLUSIONS

The purposes of this study were (1) to identify the factors which govern the rate of genetic progress by selection and (2) to develop the principles of an index, the use of which permits the attainment of maximum progress.

The aggregate genotype of an animal was expressed as a linear function of the genotypes for the economically important characters,

\[ H = a_1 G_1 + a_2 G_2 + \cdots + a_n G_n \]

where the \( a_i \)'s represent the relative importance of a unit change in their respective characters. Since \( H \) cannot be evaluated directly, it was necessary to construct a quantitative variable — an index — such that

\[ I = b_1 X_1 + b_2 X_2 + \cdots + b_n X_n \]

where the \( b_i \)'s represent regression coefficients chosen so that \( R_{ih} \) is a maximum and the \( X_i \)'s represent measurable phenotypes or linear combinations of them.

The factors which affect the rate of genetic change \( E(H) \) when selection is practiced for a correlated variable were identified as

\[ E(H) = z / p \ R_{ih} \ C_{H'} \]

where \( z \) is the height of the ordinate of the normal curve at the lowest value of \( I \) retained in a selected group and \( p \) is
the proportion of the population saved. The genetic basis for correlations between genotypes and phenotypes was investigated.

The sources of information concerning $H$ which are available in most herds were indicated, and formulae were derived for linear correlations between an animal's genotype and information from various sources based on phenotypic measurements.

The applicability of the criterion of maximum genetic progress to problems of selection was illustrated by two examples: (1) Methods of selection and (2) indirect selection for a correlated variable. From the first example it was concluded that for equally important and heritable characters which were uncorrelated, genetic progress was $\sqrt{n}$ times as fast when selections were based on an index considering $n$ characters as it is when selections were based on only one of the characters. The inefficiency of the method of "culling levels" was pointed out, particularly where the heritability or importance of the characters differed. The use of a phenotypic correlation between two characters as an indicator of the environment peculiar to the animal was illustrated by the second example.

Data taken on 12 lines of Poland-China swine which were maintained as a cooperative breeding project by the Iowa Agricultural Experiment Station and the Regional Swine Breeding
Laboratory were analyzed to obtain the constants necessary for the construction of an index having maximum efficiency. In view of investigations by Lush (1938a, 1940) the aggregate genotypic value was defined for the swine breeding program as

\[ H = \frac{1}{3} G_w + G_s + 2G_p, \]

where the \( G \)'s represent the genotypes for weight at 180 days of age, score at 225 pounds and productivity, respectively.

The information available for each pig at the time pigs were selected (about 8 to 9 months of age) consisted of:

1. Own weight
2. Own score
3. Weight of litter mates
4. Score for litter mates
5. Productivity of dam

An index based on this information was constructed (according to the principles previously developed) from the data taken during the fall of 1937 through the spring of 1940. Constants required for the formulae derived in the theoretical investigation were calculated from these data (table 14) which consisted of records taken on 1,546 pigs from 514 litters.

The standard regression coefficients (beta's) were given with \( k \), the number of pigs per litter, and \( q \) and \( M \), the coefficient of relationship between litter mates and between dam and offspring, as variables (page 85). For \( k = 5 \), the
average of the herd, and for \( q \) and \( m = .5 \) (true in panmictic populations), the index is

\[
I = .5W - .53 + .5p + .270(W - \overline{W}) + .605(S - \overline{S}),
\]

where \( W, S \) and \( p \) represent the individual weight, score and dam's productivity, respectively, \( W - \overline{W} \) and \( S - \overline{S} \) being deviations of the litter average for weight and score from the seasonal herd average.

Table 14. The Constants Necessary for the Construction of the Index.

<table>
<thead>
<tr>
<th>Phenotypic constants</th>
<th>Genotypic constants</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r_{ws} )</td>
<td>.6143</td>
</tr>
<tr>
<td>( r_{wp} )</td>
<td>-.0236</td>
</tr>
<tr>
<td>( r_{sp} )</td>
<td>-.0845</td>
</tr>
<tr>
<td>( \sigma_w )</td>
<td>31.86</td>
</tr>
<tr>
<td>( \sigma_s )</td>
<td>4.78</td>
</tr>
<tr>
<td>( \sigma_p )</td>
<td>9.71</td>
</tr>
</tbody>
</table>

The data from the pigs born in the fall of 1940 were reserved for a test of the index. The actual results of the index applied to the 1940 fall pigs were compared to those expected from theoretical considerations. Since the comparisons were generally very close, it was concluded the index
was sufficiently accurate to justify its use.

Five indexes, chosen arbitrarily, were compared to the proposed index as to the genetic progress expected by using each in a selection program (pages 94 to 96). The comparisons showed that the proposed index was most efficient.

The correlation between the aggregate genotype and the index of a pig having four litter mates was .404 for the most efficient of several indexes compared. This means that 59.6 percent of the progress possible at present levels of selection is not being realized because of the confusing effects of environment, dominance and/or epistatic deviations.

With a selection intensity of .5, the genetic progress expected per generation (as long as the genetic variances are unchanged) by selecting for the proposed index is (1) 7.76 pounds in 180-day weight, (2) 0.27 points in market score and (3) 0.25 points in productivity.
VI. LITERATURE CITED

Bartlett, M. S.

Berry, J. C.

Copeland, Lynn.

Craft, W. A.

Dobzhansky, Th.

Fisher, R. A.


Fisher, R. A., and Yates, F.

Galton, F.

Goodale, H. D.


Lush, Jay L.


Genetic correlations between relatives as influenced by limiting the number of sires. Unpublished communication. Ames, Iowa. 1938b.
Lush, Jay L.

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. 1940a.


Confidential reports to collaborators. Research Item No. 7 of the Regional Swine Breeding Laboratory. Mimeographed. Ames, Iowa. 1938.


Pearson, Karl.

Schmidt, J.
Smith, Fairfield.

Snedecor, George W.

Stonaker, R. H.

Straus, F. S.

Whatley, J. A., Jr.

Wright, S.


The effects in combination of the major color-factors of the guinea pig. Genetics 12:530-569. 1927.
Wright, S.


General, group and special size factors. Genetics 17:603-619. 1932b.


Yapp, W. W.
VII. ACKNOWLEDGMENTS

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 his guidance during the course of this study.

The author is also indebted to Professor W. G. Cochran, whose suggestions regarding statistical procedure were invaluable.