

NOTE TO USERS

This reproduction is the best copy available.

UMI[®]

A SELECTION INDEX FOR BUTTERFAT PRODUCTION
IN JERSEY CATTLE UTILIZING THE FAT YIELDS OF THE
COW AND HER RELATIVES

by

James Edward Legates

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subjects: Dairy Husbandry
Animal Breeding

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

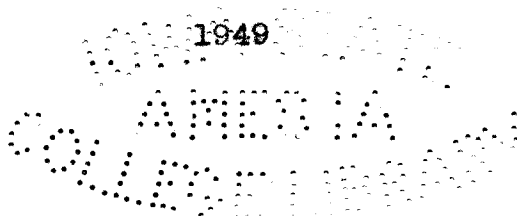
Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State College

1949



UMI Number: DP13341

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI[®]

UMI Microform DP13341

Copyright 2005 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

SF199.J5
L523s

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. REVIEW OF LITERATURE	6
A. Previous Indexes	6
B. Estimates of the Basic Statistics. . .	13
III. DATA	21
A. Source and Adjustment.	21
B. Characteristics.	26
IV. METHODS AND RESULTS.	34
A. Repeatability.	38
B. Paternal Half Sister Correlation . . .	55
C. Maternal Half Sister Correlation . . .	63
D. Heritability	68
E. The Index.	82
1. Index including cow's own production	98
2. Index for heifers or young bulls .	103
V. DISCUSSION	108
VI. SUMMARY.	118
VII. LITERATURE CITED	123
VIII. ACKNOWLEDGMENTS.	128

T9289

I. INTRODUCTION

Before man realized he could modify his animals and plants in accordance with the purpose they were to fulfill, natural selection was actively developing individuals better adapted to the existing environmental conditions. The fundamental biology which effected an inherent change from generation to generation was the same for natural selection as it was for artificial selection. Genetic¹ progress by selection is limited by the genetic variability of the trait concerned in the population under selection, the intensity of selection achieved, and the accuracy of this selection. Where simultaneous improvement in several traits is desired, possible genetic and environmental correlations between the several traits may further condition the progress that can be realized.

Lack of genetic variability is not often a serious obstacle to progress by selection except, perhaps, within strains of organisms which usually reproduce by

¹Genetic is used to express the average effect of the genes an individual possesses.

self-fertilization (Lush 1945). Low selection intensity, however, frequently may present a handicap. Intensity of selection theoretically possible is limited by the reproductive rate of the species, by whether the population is increasing, stable, or diminishing in size, and by control over losses from disease or other managerial factors. The third factor, inability to predict accurately the genotype from the phenotype, is often a major obstacle to genetic improvement. When the observed variability is largely environmental, individuals selected on the basis of their own phenotypes will have achieved most of their superiority because they experienced favorable environmental circumstances rather than because they started with superior genotypes. Development of methods to evaluate more accurately an individual's true breeding value, though its expression is veiled by environmental circumstances, has been a major topic for animal breeding investigations.

The purpose of the present study is to devise a scheme for selecting dairy cattle more accurately by using information on the production of the individual and its relatives. The technique to be proposed can be

described essentially as a combination of mass and family selection to permit maximum genetic progress. For traits which are highly heritable, mass or individual selection may be so effective that little reason remains to consider information in the pedigree, much less to wait for a progeny test, before deciding to cull or save an animal. When heritability is low, the phenotype indicates only a little about the genes the individual possesses. Under such circumstances the information on the relatives becomes comparatively more important as a means of making more rapid genetic progress. The additional information from the relatives can be used to increase the correlation between the genotype and the index (I) or basis of selection. This amounts to increasing the accuracy of predicting the individual's breeding value.

Explicitly, in this study the aid sought is in selecting for fat production in dairy cattle. The prediction equation or index is to be constructed using the fat records of the cow and her close relatives, which include her dam, her daughters, her maternal sisters, and her paternal sisters. More remote ancestors add

some further information. However, including the performance of these ancestors in the index gives more complexity and opportunity for error in computing the actual index values than is warranted by the added information. The proposed index attempts to combine the various bits of information about the performance of the cow and her close relatives in such a way that the index of an animal will be more closely correlated with its breeding value than if the information were combined in any other linear manner.

Certain fundamental statistics needed for constructing the index were required from actual data. These were calculated on an intra-herd basis; thus, the index is for selecting between cows which are all under the same general herd environment. The necessary basic statistics were:

1. Repeatability or intraclass correlation between different records of the same cow.
2. Correlation between records of paternal half sisters.
3. Correlation between records of maternal half sisters.

4. Heritability of butterfat production.

By using these statistics and the biometric relations between relatives (Wright 1921) indexes were constructed for cows which have records of their own and also have tested daughters, for heifers not in milk, and for each of these with varying numbers of maternal and paternal sisters.

II. REVIEW OF LITERATURE

A. Previous Indexes

The term index is not new in the literature on dairy cattle breeding. Numerous modifications of a "sire index" have been suggested since Hansson's proposal in 1913. Sire indexes utilize information on the sire's daughters, the mates of the sire, and possibly the herd average. They are ways of evaluating a progeny test.

Lush (1933) has already summarized the genetic implications underlying sire indexes, and in a later paper (Lush 1944) he indicated a general equation of which the several sire indexes are merely special cases. The specific problem developed in his later paper was maximizing the correlation between the sire's index value and the sire's genotype for the trait being indexed. This basically was to determine the relative emphasis to be placed on the dam's records as compared with those of the daughters. The optimum emphasis on the dam was very close to the value of the total

regression of daughter on dam. For several sets of data the unweighted average of the regression coefficients was .60 for milk yield and .55 for fat percentage. Lush further indicated that the optimum weighting of the information on the dam gave an additional 12 to 20 percent increase in genetic progress as compared with choosing the sire on the average of the daughters alone.

The study of Eldridge and Salisbury (1949) is closely related to the sire index approach, although the principle was to predict the performance of the sire's progeny. They developed an equation for predicting the performance of the daughters of young dairy sires. Fat production records on the dams, maternal half sisters, paternal half sisters, dams of the paternal half sisters, dams of the maternal half sisters, and the mates of a bull were utilized. The final prediction equation accounted for 48 percent of the total variance in the average butterfat production of a bull's daughters. The major portion, 34 percent, of the variance was accounted for by the average production of the mates of a sire. This seemingly exaggerated influence of the mates is to be expected because differences

in herd environment were ignored in calculating the regression. Daughters usually are milked in the same herd as their dams; thus, the dam accounted for most of the inter-herd variance in addition to her true genetic contribution to her daughters.

When only one trait is concerned and the information on relatives is considered, the selection index offers a way to weigh appropriately the merits of the individuals and of other members of its family. Lush (1947) has explored the results to be expected from selection solely on the basis of individuality, solely on the family average, or on the optimum combination of the two. Individual selection is more effective than family selection for characteristics which are highly heritable or when the phenotypes of the family members are highly correlated because of environmental circumstances. The family average should even receive negative attention when the phenotypic correlation between family members exceeds the genetic correlation. In such a case the family average would be more useful as an indicator of the average environment the family experienced than as an indicator of the genetic value

of the individuals in it. Individuals from a family with a high phenotypic average would be discounted since they probably experienced better than average environment. Where the phenotypic correlation exceeds the genetic correlation a phenotypically superior individual from a family with a low phenotypic average should be selected for breeding in preference to an individual phenotypically just as superior but from a family with a high phenotypic average. A combination of individual and family selection will provide greater progress than individual selection alone, although the gain will be large only if the phenotypic correlation between family members is low and the genetic correlation is high.

The first index for simultaneous selection of several traits was developed by Smith (1936) using Fisher's (1936) discriminant function technique. Smith used the discriminant function to develop a linear combination of the several traits in wheat which would best discriminate those lines with the highest aggregate genetic worth. A discriminant function was calculated for each of two years' experimental results with the same

lines of wheat. Ratios such as "straw to grain", commonly used among breeders for selection, gave less information on the genotype of the lines than did the discriminant function. In any case Smith indicated the ratios should not replace the primary observed characters.

Hazel (1943) discussed the genetic basis of selection indexes and presented the multiple correlation method of constructing selection indexes. The correlation approach is more flexible for animal populations where usually the constants needed for the index are not all based on the same number of observations. Information on relatives and information on several traits can be incorporated into the index. Where a number of traits are involved the relative economic values for the different traits, the phenotypic and genetic correlations between each pair of traits, and the genetic and phenotypic standard deviations for each trait are required. The logical basis of genetic correlations was discussed, and the procedure for estimating these correlations from parent-offspring relationships was developed in detail.

Hazel and Terrill (1946) constructed an index for selecting range Rambouillets at weaning age. Face covering, neck folds, body weight, condition, and staple length were the characteristics considered. Selection differentials for the various traits were compared for one year, where selection was on the basis of general appearance, and the succeeding year when the index was used. The authors indicated that the gain from using the index was only slight for traits in the rams where a small proportion of the lambs were saved, but was substantial for the ewe lambs where a large fraction of the lambs were saved. The authors also indicated some important practical considerations in the use of indexes for selection. Namely, a properly constructed index is advantageous in that it emphasizes highly hereditary and economically important characters and provides an objective basis for comparing individuals. The labor involved in the calculation of the index and the fact that not all selection can be based solely on the index are important limitations or disadvantages to the use of an index for selection.

Selection indexes for poultry have been developed by Panse (1946) and by Lerner et al (1947). Panse's

index involved egg weight, number of eggs, body weight, and age at first egg. The statistics on which the index was based were derived from a limited set of data; therefore the practical application of this index is somewhat limited, as the author indicates. Panse stressed the need for knowing the relative economic importance of the various traits before the selection index could be fitted most profitably to the practical situation.

The index of Lerner et al (1947) involved body weight, breast width, keel length, and shank length in New Hampshire fryers. Arbitrary relative economic values were assigned the various characters. With these indexes the increase in efficiency was from 9.9 to 14.3 percent when selection was for individuals with the highest index values as contrasted with phenotypic selection for individual characters. The need for assigning proper economic weights to each trait studied was emphasized. Furthermore, the authors indicated that only indexes involving selection for individual characters are practical until the proper economic weighting can be determined.

B. Estimates of the Basic Statistics

Gavin (1913) made one of the first attempts to determine the reliability of a single record of a cow to indicate the cow's production in a subsequent lactation. From data on 336 British Friesians and grade Shorthorns, each having five or more lactations, he found the correlation between "the revised maximum" for successive lactations to range from .39 to .58. (The revised maximum is the maximum daily yield for the lactation which is equalled or exceeded three times during the course of the lactation). Gowen (1920a, 1920b) also made early studies of the permanence of productivity of cows in a large Jersey herd. He found the average correlation between lactations of the same cow to be .52 and .54 respectively for fat percentage and for milk yield.

More recent studies on the intra-herd repeatability of fat production records standardized for age and frequency of milking are summarized in Table 1. Where the cows included in the various studies were gathered with little selection, the values for intra-herd repeatability are close to .4. Where the

Table 1. Estimates of Intra-herd Repeatability
of Fat Production Obtained by Various Workers

Data	Type of record	Repeatability estimate	Author
Over 1800 records of 552 cows in 35 Iowa DHIA ^a	365-day lactation DHIA yearly	.33 .30	Harris, Lush and Shultz (1934)
5860 records of 2316 cows in 95 Iowa DHIA herds	eight month lactation	.40	Plum (1935b)
Lifetime production records of 274 cows from 41 DHIA herds	240-day lactation 305-day lactation 365-day lactation Total lactation	.34 .34 .34 .31	Dickerson (1940)
1505 records of 301 cows in 13 Swedish herds	300-day lactation	.36	Johansson and Hansson (1940)
Records for proving 103 sires in Iowa DHIA	305-day lactation	.43	Lush, Norton and Arnold (1941)
Records for proving 209 sires in Holstein HIR ^b	365-day lactation M.E. 3x basis	.40	
Records of 454 cows in Holstein HIR with six or more lactations	365-day lactation M.E. 3x basis	.29	Berry (1945)
535 records of 226 cows in Iowa State College herd	eight month lactation	.43	Verna (1945)

^a Dairy Herd Improvement Association

^b Herd Improvement Registry

individuals studied are survivors of considerable individual selection, the true differences among cows usually have been reduced by the selection; consequently, a lower estimate of repeatability is obtained. The data studied by Dickerson (1940), Johansson and Hansson (1940), and Berry (1945) represent records of cows selected thus. Dickerson's and Johansson's data consisted of cows with five or more lactations, while Berry's data included only cows with six or more lactations. The estimate of Harris et al (1934) also may have been influenced by selection, as each cow was required to have at least two records to be included in the study.

The 365-day records are likely to be influenced more by temporary environmental conditions, such as variation in time of recurrent pregnancy, than are the 245-day or 305-day records. This may be an additional reason for the lower estimate of repeatability obtained by Harris et al. Dickerson's study, nevertheless, does not indicate the influence of such factors to be important, as repeatability was the same for the 305-day and the 365-day lactation records.

Factors based on the 70, 80, and 90 percent rule were used to adjust the data of Harris et al to a standard age basis. Much of the variance due to age differences should have been removed by this procedure, but the inaccuracies of this adjustment would leave many age differences and these would mostly be between the records of the same cow rather than between the averages of different cows. This, in turn, would give a lower estimate of repeatability than if the age differences had been more accurately corrected.

The correlation between paternal half sisters for fat production on an intra-herd basis has not been estimated from many samples of data. Plum (1934) found a correlation among paternal sisters to be .23 in a large Jersey herd including 183 daughters of 23 sires. There was a decided time trend in the production for this herd, however, and this probably contributed much to the correlation he found between the paternal sisters. In a separate study of Iowa DHIA records for 1359 cows by 611 sires, Plum (1935a) found a correlation between paternal half sisters of .196. This estimate also included some contribution from year-to-year effects within herd.

Johansson and Hansson (1940) found the intra-herd correlation between paternal half sisters to be .12 for butterfat yields corrected for age and length of calving interval. Their data included first lactation records on 1557 cows by 73 sires in 13 Swedish Red-and-White herds. The intra-herd correlation between maternal half sisters was .06 for 432 pairs. Why the maternal half sister correlation was so much less than the paternal half sister correlation was not readily apparent. The authors do point out the likelihood for a higher environmental contribution to the paternal half sister likeness due to the contemporaneity of paternal half sister.

Lush (1949) has defined heritability in the narrow sense as the fraction of the observed variance which is due to differences in the average effect of the genes the individuals possess. Table 2 presents a summary of several estimates of the heritability of intra-herd differences in fat production. All of the estimates are from parent-offspring relationships. They are given in terms of the fraction of the observed phenotypic variance in single lactation records which is additively genetic.

Table 2. Estimates of Heritability of Fat
Production Obtained by Various Workers

Data	Type of record	Heritability estimate	Author
683 daughter-dam pairs from 81 Iowa DHIA herds	eight month lac- tation	.12	Plum (1935b)
2385 daughter-dam pairs from proof of 355 sires in Iowa DHIA	365-day lactation or DHIA yearly	.25	Lush and Shultz (1936)
620 daughter-dam pairs from Iowa DHIA	365-day lactation	.13	Seath (1940)
676 daughter-dam pairs from proof of 103 sires in Iowa DHIA	365-day lactation	.28	Lush, Norton and Arnold (1941)
3010 daughter-dam pairs of 209 Holstein HIR sires	365-day lactation M.E. 3x basis	.25 and .30	
2154 daughter-dam pairs from proof of 283 sires in Iowa DHIA	305-day lactation	.17	Lush and Straus (1942)
979 daughter-dam pairs from Holstein HIR	365-day lactation M.E. 3x basis	.14	Berry (1945)
6888 daughter-dam pairs of 374 Ayrshire sires	305-day lactation	.28	Tyler and Hyatt (1947)

The estimates show considerable variability and range from .12 to .30. This variability in the estimates is a bit more striking than that exhibited in the estimates of repeatability (Figure 1). Nevertheless, the sampling errors for the estimates of heritability are comparatively larger. A number of the lower estimates Plum (1935b), Seath (1940) and Berry (1945) are based on relatively small quantities of data. Lush and Straus (1942) indicate that the five percent fiducial limits for their estimate of .17 are .03 and .31. This range includes all of the estimates given in Table 2, and the limits for the estimates on smaller amounts of data would be wider than those stated by Lush and Straus.

Where the sires have many daughters, there is a strong possibility that data for a sire may include information from more than one herd. The sire may be proven in one herd and then sold for use in a second herd or cows in calf to him may be sold into other herds where they and their daughters subsequently make records. The extent to which a sire was used in several herds perhaps is limited in most data for proving sires prior to extensive use of artificial insemination. In instances where the progeny or mates of a sire are not all

kept in the same herd, the intra-sire regression will still contain some influences of differences between herds. The estimates of heritability will be too high because of this fact. Some herd effects may have contributed to the estimates of heritability which approach .30, since all of these estimates are from intra-sire regressions in data used for proving sires.

III. DATA

A. Source and Adjustment

Data for this study were taken from the herd folders in the files of the office of the American Jersey Cattle Club, Columbus, Ohio. All herds on Consecutive Herd Improvement Registry (HIR) test during 1943 to 1947 were included. This sample involved about 75 percent of all herds that completed a year's testing during 1943 and 1944. All available completed lactation records not less than 270 days in length, from all cows in the sampled herds were used. The following information for each lactation was copied and subsequently punched on IBM cards for calculation.

1. Herd identification number (as assigned by A.J.C.C.).
2. Type classification score.
3. Cow's registration number.
4. Sire's registration number.
5. Dam's registration number.
6. Year lactation ended.
7. 305-day 2x M.E. fat production.

8. Lactation length in days if between 270 and 304 days inclusive.

The total volume of data was 23,330 lactations from 12,405 cows in 293 herds located in 42 states. The states involved and the number of herds located in each state are given below.

Alabama	1	Maryland	5
Arizona	1	Massachusetts	7
Arkansas	1	Michigan	19
California	11	Minnesota	8
Colorado	1	Missouri	19
Connecticut	5	Montana	1
Florida	2	Nebraska	2
Georgia	5	New Hampshire	2
Idaho	3	New Jersey	12
Illinois	13	New Mexico	2
Indiana	8	New York	11
Iowa	13	North Carolina	4
Kansas	6	Ohio	30
Kentucky	5	Oklahoma	1
Louisiana	1	Oregon	22
Maine	2	Pennsylvania	11

Rhode Island	2	Vermont	12
South Carolina	2	Virginia	4
Tennessee	4	Washington	11
Texas	14	West Virginia	2
Utah	3	Wisconsin	5

The advisability of standardizing fat production records for differences in age of the cow at the beginning of the lactation and for number of times milked has been adequately demonstrated. In 1943, The American Jersey Cattle Club began to correct all its fat production records to a 305-day, twice-a-day milking, mature equivalent basis (305-day 2x M.E.). This was the principal reason for taking 1943 as the beginning date for collecting the present data. American Jersey Cattle Club factors used to correct for age and for times milked are given below.

Corrections for age

<u>Age (years-months)</u>	<u>(Correction factor)</u>
Under 2	1.36
2-0 to 2-5	1.26
2-6 to 2-11	1.19
3-0 to 3-5	1.13
3-6 to 3-11	1.08

4-0 to 4-5	1.05
4-6 to 4-11	1.03
5-0	1.01
6-0	1.00
7-0	1.01
8-0	1.04
9-0	1.08
10-0 and above	1.12

Correction for frequency of milking, to be applied for
number of days cow milked 3x or 4x:

3x to 2x	.833
4x to 2x	.741

Where the lactation was actually more than 305 days in length, the portion of the last month's production needed to make exactly 305 days of production was obtained by linear interpolation for the month which contained the 305th day of the lactation. It was felt that the very short records could not be used without some adjustment, and the proper correction for these records was questionable; hence, lactations less than 270 days were not used. Short lactations result partly because of inherent lack of persistency in some cows (Bonnier 1935); therefore, adjusting such lactations to

a standard length would also eliminate some real differences in productivity. Lush and Shultz (1936) estimated that less than 10 percent of the variance in lactation length seemed to be hereditary in the simple additive manner, and only one-seventh seemed attributable to general herd managemental policy. In most cases a justifiable clue was not available to indicate why the cow did not complete a ten-month lactation. For these reasons, short lactations were eliminated and all lactations of 270 to 304 days were used with no adjustment for length.

Use of the 305-day lactation record also eliminates most of the variation in production which could be attributed to variation in the time of recurrent pregnancy. Dickerson (1940) found that 23.6 percent of the variance in total lactation butterfat yields was associated with variations in length of calving interval. When 305-day fat yield was considered, he found that only 4.8 percent of the variance in production could be attributed to differences in length of calving intervals.

B. Characteristics

The average fat production on a 305-day 2x M.E. basis for all records was 429.4 pounds. This value cannot be compared with the average actual production of all cows on HIR test since the age composition of the herds was not available to permit a reasonable adjustment of the latter for age. Average production for all herds is reported by the American Jersey Cattle Club in terms of the actual production and includes the production of all cows in the herd for ten months or more during the testing year. Again the elimination of lactations which were less than 270 days in length and using only the first 305 days of longer lactation for this study is different from the procedure followed by the club in computing HIR herd averages.

Information concerning the probable intensity of selection for production in these herds is only sketchy and inconclusive. The average production of the 3363 dams that made lactations during 1943-47 was 425.5 pounds whereas the average production of their 4764 daughters was 426.7 pounds. The slight difference

between the daughters and the dams can probably be explained by the fact that the dams in the higher producing herds had more daughters in production than dams in lower producing herds. The average for the daughters would be made up proportionately of more cows from the higher producing herds than the average of the dams.

Table 3. Average Production According to the Number of Lactations per Cow

No. lactations per cow	Ave. prod.	No. cows
1	406	5870
2	423	3345
3	440	2160
4	454	862
5	465	166

Although the above figures do not indicate selection as regards the production of the dams of cows kept in these herds, there is some slight indication that breeders were weeding out some low producers. Table 3 shows the average production of all cows according to the number of lactations included in the study.

Since the data include only a five-year period,

the cows with one lactation include some old cows completing their last record in 1943 and some young cows completing their first lactation during 1947. Whether the culling was willful or accidental could not be determined from the information available. There is, nonetheless, a definite trend for the cows with more records to have a higher average production. A portion of this trend can possibly be attributed to the fact that cows in the higher producing herds, on the average, have more records than cows in lower producing herds. However, since the number of records per cow does not vary greatly and only a few of the records involved were made in the very low or very high producing herds, it seems unlikely that this could account for a major part of the trend.

The pooled intra-herd variance was apportioned in the various aspects of the study. This has been a common procedure in the analysis of dairy data collected from several breeders' herds. The question as to whether the several intra-herd variances pooled were homogeneous apparently has not always been considered. Bartlett's test as described by Snedecor (1946 p. 251) was applied

to the intra-herd, intra-year variance in the present data. A highly significant χ^2 value was obtained, indicating that the variances probably were not all sampled from the same population.

Considering the data from another aspect, the correlation between the mean production of the herd and the intra-herd intra-year variance was $.46 \pm .05$. Because of this and to investigate further the relation of the mean to other aspects of the study, the herds were grouped according to the production levels shown in Table 4. Even in this case highly significant χ^2 values were obtained when Bartlett's test was applied within each production group. The conventional recommendation to group the data to obtain segments with homogeneous variance offered little toward solution. Satterthwaite (1946) has indicated that by the use of components of variance even differences between sample means may be tested, ignoring any assumptions regarding homogeneity of their variances. Since estimates of components of variance, rather than precise tests of significance, were desired from these data, it seemed that the heterogeneity should not bias estimates of the components.

The scatter diagram of the mean production plotted with respect to the intra-herd intra-year variance is shown in Figure 1. The possibility of a slight curvilinearity of this relation also can be noted. Johansson and Hansson (1940) and Bonnier (1946) have observed the increased intra-herd variability as the mean production rises in data from Scandinavian herds. Their explanation is that the true differences among cows are permitted full expression only when the cattle are well managed. Johansson and Hansson express the opinion that since true productive capacity of cows can be expressed only in an optimum environment, the correlation between daughter and dam should be larger in the higher producing herds. A preliminary look at this point in these data does not indicate that this relation exists. Gowen (1920a) has shown that a similar increase in variability of production records exists as the cow approaches maturity. The relation between the mean production and age was a logarithmic function which suggested multiplicative action of genes influencing production. This type of gene action might also explain the increase in intra-herd variability as the herd mean production increased.

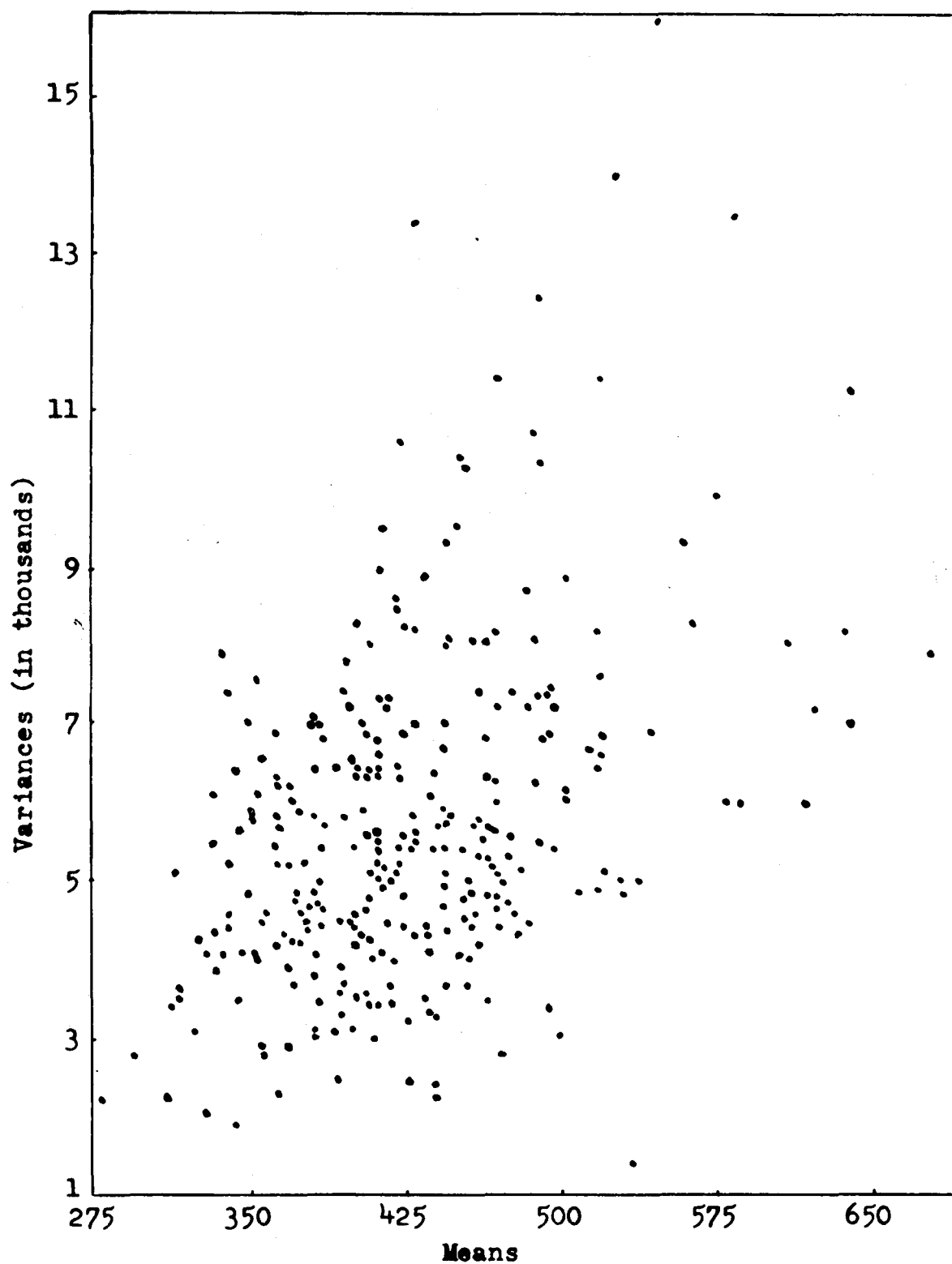


Figure 1. Scatter Diagram Showing the Relation Between the Herd Mean Fat Yield and the Intra-herd Intra-year Variance for 293 Herds.

Table 4. shows the same relationship between the mean production and intra-herd variability in another light. The pooled intra-herd intra-year coefficient of variation is given for each of the nine production groups. A slight trend toward a decrease in the coefficient of variation as the mean production increases is evident. The standard deviation increased but not in direct proportion to the increase in the mean. A marked increase in the standard deviation and a slight decrease in the coefficient of variation as the mean rises can be construed to mean that many of the factors, both genetic and environmental, which influence fat production do so multiplicatively (i.e. on a percentage basis).

Table 4. Pooled Intra-herd Intra-year Coefficient
of Variation for the Nine Production Groups

Prod. range for herd means	No. herds	No. records	Mean (\bar{x})	Std. dev. (σ)	C. of variation ($\frac{\sigma}{\bar{x}}$)
277-330	12	880	313	61.1	.195
331-360	29	2291	346	71.8	.207
361-390	44	3369	377	71.0	.188
391-420	65	4915	407	76.3	.188
421-450	49	3558	436	77.3	.177
451-480	45	4328	462	77.2	.167
481-510	20	1488	492	87.3	.177
511-540	13	1045	518	80.9	.156
541-675	16	1456	590	93.4	.158

43330

IV. METHODS AND RESULTS

Wright (1921, 1935) demonstrated that only the average effects of genes contribute to permanent gain from selection. Gene effects ascribable to dominance and epistatic deviations act much as temporary environment in that they prevent the individual's phenotype from accurately indicating the genes the individual possesses. A convenient mathematical concept to express a trait in a discussion of improvement by selection is

$$X = G + U.$$

The sum of the average effect of the genes is designated by G^1 , and U represents the combined effects of environment and dominance and epistatic deviations. With the above definition and the assumption that the G values are uncorrelated with U , the observed phenotypic variance in X is

$$\sigma^2_X = \sigma^2_G + \sigma^2_U.$$

The total hereditary variance in a given trait can then

¹Throughout the discussion G will be referred to synonymously as breeding value, average genic value, or genetic merit.

be expressed

$$\sigma^2_H = \sigma^2_G + \sigma^2_D + \sigma^2_I$$

where σ^2_D represents the dominance deviations from the additive scheme or the interaction among alleles, and σ^2_I refers to the epistatic deviations from the additive definition or the interactions among non-allelic genes. Where a true interaction between H and E (environmental circumstances) exists a portion of the observed variance will be referred to as σ^2_{HE} . The proper apportioning of this variance to heredity or environment is yet an unsolved question, as it is a joint effect of variation in both H and E.

Again assuming the model introduced above for a character X and making the further assumption that X and G are normally distributed, the relative efficiencies of various selection programs can be compared. Smith (1936) and Hazel and Lush (1943) have demonstrated that if the selection intensity for a given population is a constant the genetic advance or progress from selection is proportional to the correlation between G and X. Under the simple situation where only one trait is considered the expected gain from truncation selection is

$$E(\hat{G}-\bar{G}) = r_{GX} \frac{Z}{P} \sigma_G$$

where Z is the height of the ordinate of the normal curve at the lowest value of X saved and P is the fraction of the original population permitted to have offspring.

For a single character r_{GX} is equal to $\frac{\sigma_G}{\sigma_X}$ which is the square root of heritability in the narrow sense. If selection can be based on an index or a correlated variable (I) such that $r_{GI} > r_{GX}$, the progress by selection will be enhanced. The biology of the population is not affected, since the same proportion of individuals presumably would need be saved regardless of the basis of selection.

Hazel and Lush (1943) explored the consequences of selection on the basis of total score, independent culling levels for each trait, and tandem selection. Tandem selection, which means selection for one trait at a time until a given level of improvement is reached followed by subsequent selection for another trait, was the least efficient. Total score or index selection was the most efficient of the methods considered. When a large number of characters are to be selected simultaneously, the selection index approach was indicated to

be even more efficient comparatively than when only one or a few characters were being selected.

To construct the index in the present study the following statistics were required from the data:

1. Repeatability or intra-class correlation between different records of the same cow.
2. Correlation between records of maternal half sisters.
3. Correlation between records of paternal half sisters.
4. Heritability of fat production.

These basic statistics and the path coefficient approach to the biometrical relations among relatives (Wright 1921) were employed to construct an index. The first three statistics were estimated as intra-class correlations; whereby, the observed variance is separated into its various components depending on the mathematical model used to describe the biological situation. Fisher first indicated this separation of the variance in his early text. The idea has been extended and restated by Winsor and Clark (1940) and more recently by Crump (1946). Hetzer, Dickerson, and Zeller (1944) also have given

details for computing coefficients for the expected numbers of various components in the mean squares, where the number of observations per subclass varies. Heritability was estimated by doubling the intra-herd regression of daughter on dam.

A. Repeatability

The linear mathematical model used to describe each lactation was:

$$Y_{ijk} = \mu + h_i + y_{ij} + c_{ik} + e_{ijk}$$

where:

Y_{ijk} is the 305-day 2x M.E. fat yield of the k th cow in the i th herd and in the j th year.

μ is the effect common to all observations.

h_i is the effect ascribable to the i th herd and

$$E(h_i)^2 = \sigma_h^2 = H.$$

y_{ij} is the effect ascribable to the j th year in the i th herd and $E(y_{ij})^2 = \sigma_y^2 = Y.$

$E()$ is used to indicate the expected or average value of the term in the parenthesis.

c_{ik} is the effect of the k th cow in the i th herd

$$\text{and } E(c_{ik})^2 = \sigma_c^2 = c$$

e_{ijk} is the random error in evaluating the cow's production, and the effect of temporary environmental influences associated with the expression of each lactation and

$$E(e_{ijk})^2 = \sigma_e^2 = e$$

The present problem is to estimate the variances or variance components ascribable to the various effects. In such a case no assumption need be made concerning the form of the distribution of the several effects. What must be assumed is that the effects which make up each observation are random variables independently distributed about means zero (Crump 1946).

The herd effect (h_i) is defined to be the sum of all influences which make the mean production of one herd differ from the mean production of a second herd. Such differences in production may result from variation in the environments or managerial regimes from herd to herd and also from true differences in the intrinsic merit of the cows in the several herds. No

attempt is made in the present study to separate the intra-herd variance into a portion resulting from intra-herd environmental differences and a remainder resulting from differences in the real producing ability of the cows in the several herds. However, an estimate of what fraction of the herd component is due to differences in the mean breeding values of the individuals in different herds would be desirable for making breeding plans most effective, especially where selections must be made between animals in different herds.

The year effect (y_{1j}) includes the several environmental influences which cause the yearly mean production for the herd to fluctuate about the general mean of that herd. Variation in the quantity and quality of feed available from year to year, or managemental practices which prevailed in one or more of the years, but not in all years, are examples of influences considered to be in y_{1j} .

All things which make permanent differences in the producing abilities of the several cows in the same herd are defined as the cow effect (c_{1k}). The variance

associated with this effect (C) includes the sum of the variances due to σ_G^2 , σ_D^2 , σ_I^2 , and σ_{PE}^2 . The variance denoted σ_{PE}^2 is the result of permanent environmental influences which affected alike all the records of the same cow but were different for other cows. Its definition is somewhat arbitrary. For example, some effects may be influential for two consecutive lactations of a cow but not for her other lactations. If only those two were in the data, this effect would be "permanent" but, if several additional lactations of the same cow are in the data, this effect permanent to two of the lactations would now be defined largely as temporary. Berry (1945) found these partially permanent influences to be of minor importance, and they probably contributed from .03 to .09 to the intra-herd correlation between records of the same cow in his data.

The error effect, (e_{ijk}) includes true, presumably random, error in evaluating the yield of the cow and also the influence of temporary environmental circumstances which genuinely varied for the several records of the same cow. The component E indicates the importance of the unclassified and uncontrolled sources

of variation in causing the observed variance of yields within a given herd.

Intra-herd repeatability, after accounting for the variance due to differences between years within herd, is defined as $\frac{C}{C + E}$. This ratio represents the fraction of the observed variance in records made in a given herd, but corrected for yearly differences, which is due to intrinsic differences in producing ability among cows in that herd. Repeatability is the intra-herd correlation among the several records of the same cow, if the effect of Y were first removed. The logic in describing this ratio of variance components as a correlation is apparent when we consider that among records of the same cow the variance C vanishes, and only the variance E remains. Removal of variance C, by holding the cow constant, represents the degree to which observations on a given cow are alike when compared with observed differences present in a population of records from many cows in one herd.

As indicated in the mathematical model given on page 33 an attempt was made to free the estimates of C and E from contributions due to yearly herd variations

(Y). Since few of the cows in this study had records in all five years, the differences between the yearly means for each herd did not reflect only the yearly environmental changes.

The least squares method of fitting constants for each cow can furnish an estimate of year effects, where each herd is considered an analytical unit. But the use of the least squares procedure gives a biased estimate of the true year effects because cows with two or more records, from which the year effect is estimated, are in practically all cases a selected group. Thus, the cells which are filled in the two-way cow x year table are not a random sample of all possible cells. Cows completing a good record are more likely to remain in the herd another year than are cows completing a low record. These cows are chosen in part because they truly possess above average inherent producing ability. However, since fat production is influenced greatly by environmental circumstances, the selected group of cows generally had better than average temporary environment in the year when they made the records on which they were selected. Assuming these

temporary environmental conditions to be random, the average environmental influences for the cows saved could hardly be expected to be as favorable the second year as they were for the year in which the cows were selected. Because of this the least squares procedure may indicate the environmental circumstances to have been getting poorer and poorer; whereas, in truth they may have been constant from year to year.

Henderson (1949) has developed a maximum likelihood procedure for estimating the yearly effects and for separating the year-to-year changes in herd environment from the changes in real producing ability of the cows in the herd. This procedure takes into account the fact that the observations available to estimate year-to-year changes are a selected sample and that the successive records of the same cow are incompletely repeatable. For the present study, however, the estimation of components of variance, rather than finding the true magnitude of the effects, was the main problem. These components can be estimated by several procedures. Information on methods of evaluating the relative efficiencies of these procedures are uncertain. The method

chosen was to compute the analysis of variance as though the data were orthogonal and to suppose that the mean squares found in the sample actually are the expected or average values. This procedure yields unbiased estimates of the components ($E(\theta) = \theta$) but the validity of the values obtained are dependent as in all cases, upon the sufficiency of the basic model.

The analysis in Table 5, employed to estimate C and E free from the influence of Y, was suggested by Dr. Jay L. Lush. Its validity depends on c_{ik} and y_{ij} being uncorrelated, as is inherent in the definition of Y. Estimates of Y, C, and E were obtained by simultaneous solution of equations, using the four sample mean squares : among years within herds, within year, among cow within herd, and within cow. The estimates of Y, C, and E were then substituted, as is conventional, into the equation for the mean square among herds to estimate the herd component H.

The lactation records at the offices of the A.J.C.C. are recorded according to the year in which the lactation was completed. Thus, in a few instances a cow was credited with completing two 305-day lactation

Table 5. Analysis of Variance for the
Influence of Herd, Year Within Herd, and Cow

Source	d/f	Mean square	Expected composition of mean square
Total	23,329	10,498	E + .999 C + .999 Y + .995 H
Among herds	292	342,207	E + 2.342 C + 19.177 Y + 79.482 H
Among years within herd	1,062	13,527	E + .630 C + 16.669 Y
Within year	21,975	5,944	E + C
Among cow within herd	12,112	8,607	E + .560 Y + 1.870 C
Within cow	10,925 23,637	3,729	E + Y

Components

H = 4060

Y = 515

C = 2730

E = 3214

records in one year. One of the records would necessarily have to be made largely during the previous year. In computing the expected composition of the mean squares in Table 5 the simplifying assumption was made that a cow had only one record in a given year.

Table 6. Components of Variance as Fractions of Total and Intra-herd Variance

Component	Estimate of component	Fraction of variance represented	
		Tot. (H+Y+C+E)	Intra-herd (Y+C+E)
H	4060	.386	
Y	515	.049	.080
C	2730	.259	.422
E	3214	.306	.498

16519

The estimated components are given in Table 6 as fractions of the total and as fractions of the intra-herd variances they represent. Ratios of these variance components furnish estimates of several intraclass correlations. For example, there is a correlation of .386 between single records of herd mates when a population of cows from many herds is considered. The principal estimate from this analysis needed for the index is that for repeatability. Among records appropriately adjusted to

remove year-to-year environmental influences intra-herd repeatability, $\frac{C}{C+E}$, is .459. The practical usefulness of this figure cannot be realized until lactation records can be accurately adjusted for yearly herd environmental changes.

When those yearly differences are not removed from the data, intra-herd repeatability would be $\frac{C}{C+E+Y}$, which equals .422, if none of the cows being compared were contemporary. Since in actual data many of the cows will be at least partly contemporary, the Y component will be only partially represented in the mean square "among cows within herd" although fully represented in the mean square "within cow". This will lead to a slight underestimate when C is computed merely by subtracting the mean square within cow from that among cows within herd, and dividing the remainder by the coefficient of C in the latter mean square. Using the estimates of C and E obtained in this manner gives an intra-herd repeatability of .412. If herd differences were ignored, as in many of the early studies, repeatability would be nearly .65. The importance of first removing herd difference becomes strikingly apparent

$$\begin{array}{r}
 2609 \\
 3219 \\
 \hline
 2609 + 3219 = 5828
 \end{array}$$

when this value is contrasted with intra-herd repeatability from the same data of only a little over .4.

According to the z transformation of Fisher (1946) the 90 percent confidence interval for this estimate, .412, is .399 to .422, lending support to the idea that the true parameter value is not greatly different from the estimate obtained from these data. The value, .412, was selected for use in constructing the index, since it more closely represents what can be realized in a practical breeding program where an accurate adjustment for yearly effects is not available.

The herd component accounts for 39 percent of the total variance. Although the herds in this study were limited to those on HIR test for at least four years, this restriction apparently had little or no tendency to secure a group more uniform in mean production than a truly random sample of all HIR herds probably would have been. The herd mean production ranged from 277 to 675 pounds. These herds were from all sections of the United States. Probably they represent a wide range of systems of management followed.

The present estimate of the importance of herd differences agrees closely with a number of estimates from analyses of Iowa DHIA records. Harris et al (1934), Plum (1935b), Lush and Shultz (1936), and Lush and Straus (1942) found herd differences accounted for approximately 33, 35, 33, and 30 percent respectively of the total variance in corrected lactation records.

Dickerson (1940) found herd differences accounted for only 17 percent of the total variance in age corrected 305-day fat records. He attributed the smaller influence of herd differences to the selected nature of his data, which included only cows with at least five records each. The herd component also can be calculated from Bonnier's (1946) data on 406 herds of the Swedish Red-and-White breed in one county in Sweden. In those data the herd component was approximately 30 percent of the total variance in yearly records.

Intra-herd year-to-year differences in these data accounted for about eight percent of the intra-herd variance and five percent of the total variance in individual lactation records. In Iowa DHIA data Plum (1935b)

found that year-to-year differences accounted for six to seven percent of the total variance and slightly more than ten percent of the intra-herd variance. That is somewhat more than in the present study. Johansson and Hansson (1940) found that differences among years within herds accounted for between five and six percent of the total variance. They express the opinion that in well managed herds yearly fluctuations about the general herd trend are a minor source of variation.

In the present data the component Y as defined on page 40 includes any general year effect over the entire sample of data plus what might be considered true herd x year interaction. The general year effect might result from nation-wide price changes or feed shortages which would cause dairymen generally to provide management more favorable to high production, perhaps in 1946 than in 1944. The herd x year interaction would result from herd environment improving in some herds in the same year it deteriorated in other herds.

Table 7 shows the analysis of variance and the estimates for the components attributable to (\underline{y}_{1j})

Table 7. Analysis of Variance for Years,
Herds, and Herd x Year Interaction

Source	d/f	Mean square	Expected composition of mean square		
Total	23,329	10,498	E + .999C + .999HY +	.995H +	.792Y
Among herds	292	342,207	E + 2.342C + 19.177HY +	79.482H +	2.649Y
Among years	4	286,487	E + .642C + 28.213HY +	5.064H +	4618.755Y
Herds x years	1,058	12,495	E + .630C + 16.625HY -	.019H -	.731Y
Within year	21,975	5,944	E +	C	
Components					
H = 4071	Y = 55.		HY = 462	C = 2730	E = 3214

the average year effect, (hy_{1j}) the herd x year interaction, and (h_1) the herd effect. For nonorthogonal data with multiple classification the main effects cannot be separated without entanglement. Although the lack of orthogonality does exist, unbiased estimates of the components should be obtained from the expected values of the respective mean squares. Estimates for E and C were available from the previous analysis. These values were substituted into the equation for the mean square among years, among herds, and for the herd x year interaction. Estimates of \bar{Y} , HY, and H were obtained by simultaneous solution of the three equations involving these components. As indicated by the estimates, \bar{Y} makes up only a little over 10 percent of Y. That is nearly all of Y comes from HY. It may be noted that the estimate of H is slightly different from the estimate obtained in Table 5. This may have resulted from an automatic although slight change in the definition of H in the second analysis as a function of the nonorthogonality of the data. Since the difference in the two estimates of H is one-fourth of one percent of the absolute value of H, the discrepancy seems of no practical importance.

Corroborating evidence as to the unimportance of a general time trend in these data is offered by the information in Table 8. Although a slight upward trend is evident, most of the variance associated with the general year effect is in the contrast between the means

Table 8. Average Production for all Lactations by Year Lactation was Completed

Year	Ave. prod. (lbs)	No. records
1943	421	3313
1944	420	4780
1945	433	5457
1946	432	5849
1947	439	3931

for 1943 and 1944, and the means for the last three years. The reason for the decided rise in average production from 1944 to 1945 is not evident. However, the variation in the yearly means given in Table 8 involves some year-to-year change in the cows in the several herds as well as the general year influence.

B. Paternal Half Sister Correlation

To compute the correlation among paternal half sisters, the portion of the intra-herd variance associated with the sire had to be ascertained. Since paternal half sisters have a common sire, the variance within groups of paternal half sisters would be reduced by an amount associated with the true difference among sires.

The model chosen for determining the component of variance attributable to the sire is similar to the one indicated on page 38.

$$Y_{ijkl} = \mu + h_i + Y_{ij} + s_{ik} + P_{ikl} + e_{ijkl}.$$

The difference in the two expressions is that the above equation assumes the separation of the variance attributable to the cow effect (c_{ik}) into two additive components. The first, S , is associated with the sire effect (s_{ik}); and the second, P , is associated with (p_{ikl}) real differences among cows which have a common sire.

Table 9 gives the analysis of variance made to determine the paternal half sister correlation. Some

Table 9. Analysis of Variance Among Sires,
Paternal Half Sisters, and Full Sisters

Source	d/f	Mean square	Expected composition of mean square
Total	23,329	10,498	$E + .999Y + .999P + .999S + .995H$
Among herds	292	342,207	$E + 19.177Y + 2.342P + 14.214S + 79.482H$
Among sires within herds	4,041	11,083	$E + .775Y + 1.898P + 4.741S$
Among paternal half sisters	7,123	7,311	$E + .449Y + 1.850P$
Among full sisters	948	7,797	$E + .474Y + 1.897F^a$
Within cow	10,925	3,729	$E + Y$

Components

H = 3947 S = 741 P = 2090 F = 2287 E = 3214 Y = 515

^a F is used to designate the variance due to real differences between full sisters.

of the comparisons among daughters of the same sire were comparisons between full sisters. Full sisters are expected to be more alike than half sisters. If this is the actual situation, ignoring the full sisters would bias the estimate of the paternal half sister correlation toward being too large. To avoid this, the sum of squares associated with differences among full sisters was subtracted from the sum of squares for differences among cows by the same sire. The remainder was a sum of squares associated only with differences among paternal half sisters.

Estimates of Y and E were available from Table 5. By using these values the other components could be estimated. The correlation between paternal half sisters, all kept in the same herd with their records appropriately adjusted for the influence of yearly herd environmental variations, is $.123 \left(\frac{S}{S+P+E} \right)$. If the correlation is evaluated ignoring the Y component, a value of .120 is obtained. The components E and S are increased and P is decreased, but the ratio is hardly affected when Y is not considered. In a population in which Y is a much larger fraction of the total variance than it is in the present

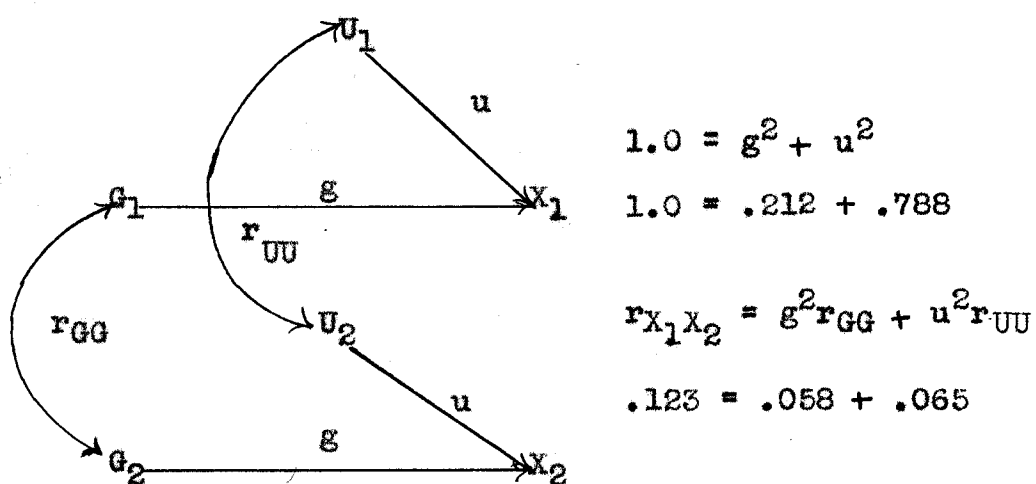
data, it would have more influence on the correlation.

The distribution of intraclass correlations involving more than two components is not known exactly. Satterthwaite (1946) has given an approximate distribution of variance components from which the correlations are derived. However, the sampling errors of the components making up the correlation are correlated. Hence, knowledge of the sampling errors of the components does not give a good approach to making a probability statement concerning the correlation. But significance of the correlation can be tested by the F test of the mean squares involved in estimating the components. The paternal half sister correlation from these data is significant beyond the $P(.01)$ level.

An estimate of heritability can be obtained in random breeding populations by multiplying the half sister correlation by four, after first correcting for the environmental contribution. Evaluating this environmental influence, however, is difficult. Rarely can it be determined directly. If the estimate of heritability (.212) obtained from the intra-herd daughter-dam regression (page 79) can be assumed to represent σ_G^2 / σ_X^2 , an

estimate can be made of the environmental contribution to the paternal half sister correlation.

The diagram below illustrates the relations between the breeding values (G) and the fat yields (X) of two paternal half sisters. Each is expressed as a



deviation from their herd means, and the fat yields are adjusted for yearly herd environmental influences.

Assuming the genic correlation between mates of a sire to be .1 and the genic correlation between the sire and dam (m) to be zero, r_{GG} equals .275. where heritability is .212 as estimated from the intra-herd daughter-dam regression. If the above assumptions are reasonable, the environmental contribution to the likeness between

paternal half sisters is of about the same magnitude as the contribution from the average effect of similar genes which they possess. Since the variance Y was removed before computing the paternal half sister correlation, it would appear that most of the environmental contribution to the correlation should already have been discounted. However, when Y is ignored in computing the correlation the value is, .120, which is practically the same as the value obtained when Y was removed first. If we were to assume that all of the environmental correlation between paternal half sisters is caused by Y, we would obtain an estimate of nearly .5 for intra-herd heritability. This can be contrasted with the value .21 obtained from much of the same data by the daughter-dam regression. Apparently considerable environmental correlation remains in the paternal half sister correlation although the physical cause or basis for it is not clear.

The existence of an environmental correlation between paternal half sisters limits the efficiency of the progeny average as an indicator of the sire's genotype. If each daughter has one record, the correlation between the genotype of the sire and the average

production of his \underline{n} daughters, can be expressed as:

$$r_{G_s \bar{X}} = g/2 \sqrt{\frac{n}{1 + (n-1)t_{XX}}} \quad (\text{random mating assumed}).$$

The phenotypic correlation between paternal half sisters (t_{XX}) is .123. This correlation would be $g^2 r_{GG}$ or .058 if paternal half sisters were alike only because of the genes they have in common and if heritability is estimated correctly from the daughter-dam regression. Where a sire has five daughters, $r_{G_s \bar{X}}$ would be .411 with t_{XX} equal to .123, but .452 if t_{XX} equals .058. Each additional daughter furnishes proportionately less information about the breeding value of the sire, especially if the environmental correlation between the daughters is increased. In the limit where the number of daughters becomes infinitely large, $r_{G_s \bar{X}}$ would approach 1.0 if the correlation between paternal half sisters was only $g^2/4$. However, the limit in the present case is $\sqrt{\frac{g^2/4}{g^2 r_{GG} + u^2 r_{UU}}}$ or .656. The environmental correlation between paternal half sisters and the assumption that the mates of the sire were related ($c = .1$) has reduced the value toward which $r_{G_s \bar{X}}$ asymptotes as \underline{n} becomes large.

In Table 9 the variance between full sisters (F)

is larger than the variance between paternal half sisters (P). Full sisters have both sire and dam in common and should be alike in half the genes they possess. Only if the environmental correlation between paternal half sisters is much larger than the environmental correlation between full sisters would F be expected to be greater than P. The reason for the greater variance between full sisters in these data is not clear. However, the full sisters appear to be a selected group of cows in that their average production is 447 pounds as contrasted to 429 pounds for all cows. Herds with higher mean production also tend to have greater intra-herd variability. In so far as this might be true, the variance (E') between records of the same cow may be somewhat larger for full sisters than the pooled estimate from all cows. If the value of E' in the mean square between full sisters were larger than the value substituted (the pooled estimate), the estimate of F would contain some contribution from the error variance and would in effect be too large.

C. Maternal Half Sister Correlation

In the preceding section the estimate of the variance associated with real differences among sires was used to estimate the paternal half sister correlation. A similar procedure involving the progeny classified according to their dam, whereby the correlation between maternal half sisters can be estimated, is followed in this section. The equation or mathematical model assumed for each lactation was:

$$Y_{ijkl} = \mu + h_i + y_{ij} + d_{ik} + m_{ikl} + e_{ijkl}.$$

The component of variance associated with the effect of the dam (d_{ik}) is defined as D. This component represents the variance resulting jointly from the genetic contribution of the dam and, of course, other conditions which were alike for all daughters of a dam but differed for other herd mates. If a maternal influence on production resulting from intra-uterine environmental effects is truly present, it would contribute to D also. This maternal effect, representing an environmental influence of the dam on the offspring, might well be appreciable for traits such as weaning weight of pigs

or beef calves. However, it seems unlikely that such an influence is important for a trait like fat production which is not expressed until at least two or three years after birth.

The component of variance associated with m_{1kl} is designated M. It represents the variance due to real differences among cows all from the same dam. In a random breeding population the variance of true differences between maternal half sisters should include

$$3\sigma_G^2/4 + \sigma_D^2 + \text{most of } \sigma_I^2 + \text{most of } \sigma_{PE}^2.$$

Table 10 gives the analysis of variance used to estimate D and M. The sum of squares for differences between full sisters was subtracted from the sum of squares among cows within dam. The remainder would be associated with differences among maternal half sisters freed from the comparisons between full sisters. The values of the components E and Y were taken from Table 5 to solve the equations for the mean squares among maternal half sisters and among dams within herd.

The correlation between maternal half sisters kept in the same herd, with their records appropriately

Table 10. Analysis of Variance Among Dams
Maternal Half Sisters and Full Sisters

Source	d/f	Mean square	Expected composition of mean square
Total	23,329	10,498	$E + .999Y + .999M + .999D + .995H$
Among herds	292	342,207	$E + 19.177Y + 2.342M + 3.384D + 79.482H$
Among dams within herds	9,621	8,834	$E + .570Y + 1.871M + 2.322D$
Among maternal half sisters	1,543	7,695	$E + .549Y + 1.840M$
Among full sisters	948	7,797	$E + .474Y + 1.897F$
Within cow	10,925	3,729	$E + Y$

Components

H = 4054 D = 455 M = 2282 F = 2287 E = 3214 Y = 515

corrected first for year-to-year environmental influence, is estimated by the ratio $\frac{D}{D + M + E}$. This correlation is .077 in these data. Nevertheless, this seems to be a slight underestimate of the actual similarity between maternal half sisters in these data. In this study 8013 of the total of 9914 dams had only one daughter. With only one or two sires in service at a time a number of the intra-herd comparisons between dams involve contrasts between paternal half sisters. This would make the mean square among dams too small, since there would not be a sire difference in all the comparisons between dams. The component M as computed from the mean square among maternal half sisters includes a full sire component (S). When M is substituted in the equation for the mean square among dams within herd, too much is subtracted, thus leaving an underestimate of D.

An estimate of S was available from Table 9. This can be used to indicate the reduction in the mean square among dams for those cases where the dams are represented by progeny which are paternal half sisters. In another way of thinking, S would be the reduction in the mean

square as a result of holding the sire constant. From sampling a few of the herds it appeared that about .15 of the intra-herd comparisons between dams involved comparisons between paternal half sisters. The average value of M in the mean square among dams would be (.15) (S) smaller than the estimate of M obtained directly from the comparison of maternal half sisters.

If the D component is corrected thus for the influence of the common sire in some of the comparisons, the correlation between maternal half sisters rises to .090. The estimate of the D component then becomes 556. If the Y component of variance is ignored in estimating the correlation, a value of .086 is obtained. Again the values of the individual components are influenced much more by ignoring Y than is the ratio needed to evaluate the correlation.

In these data estimates of variance due to real differences between maternal half sisters (M), paternal half sisters (P), and full sisters (F) could be made. Theoretically, these should furnish an estimate of variance attributable to maternal effects and indicate the possible importance of variance due to dominance

deviations. Subtracting M from P should give an estimate of variance associated with maternal influence. Since P is smaller than M, presumably as a result of stronger environmental correlation between paternal half sisters, the estimate is negative.

Variance between full sisters should be less than that between paternal half sisters by a quantity equal to the dam component (D) plus one-quarter of the variance due to dominance deviations. Full sisters are expected to be alike for one-quarter of the gene pairs which would be expected to contribute to dominance deviations between half sisters or unrelated individuals (Fisher 1918). The difference between M and $F + P$ also should furnish an estimate of the importance of dominance deviations. Neither of the two alternatives are useful in these data, since the estimate of F was much larger than would be expected theoretically.

D. Heritability

Heritability was estimated by doubling the intra-herd regression of daughter on dam. Lush (1949) compared

a number of methods for estimating heritability of characteristics of farm livestock. The intra-sire regression of daughter on dam has been a most convenient scheme for estimating heritability of fat production in dairy cattle. Lush (1949) indicates several advantages of the intra-sire regression to estimate heritability. Most of the herd differences, presumably largely environmental, would be removed in the differences between sires. Complete removal would result, however, only if all of a sire's daughters and mates are born and kept in the same herd. Where data are collected over a long period of time, trends in management may be a potent contributor to the resemblance between daughter and dam on an intra-herd basis. The progeny of a sire are usually contemporaries and most of the variance attributable to managemental trends would go into the differences among sires.

An argument for using the regression instead of the correlation, is that the dams are usually a selected group. The dams should represent a sample with lower variability than the population from which they were chosen. Such selection would tend to lower the correlation

between daughter and dam, but selection of the independent variate would not bias the regression (Eisenhart 1939).

Lush (1948) points out a further caution that if selection of the dams was not solely on the basis of their own phenotypes for the trait being considered, the regression may be biased toward smallness. The magnitude of this possible bias is conditioned by the amount of attention given the auxiliary information and by the intrinsic relation between the other items and the genic value for the character under consideration.

For these data the herds in which both the daughters and the dams made their records could be easily identified and the data were collected within five years during which there was not an important time trend. Thus, the intra-herd rather than the intra-sire regression was computed in the present study. The average of all available records of the daughter and the dam was used to compute the daughter-dam regression. As pointed out by Lush (1942) the variance of the dams with an average of n records is only $\frac{1 + (n-1)v}{n}$ of the variance calculated

for a random single record of each dam (v is the repeatability of records of the same cow). Thus the regression computed by using the average of all records of the dam would be higher than that obtained using a random record of the dam.

The linear model used to express a lactation fat yield of the dam is:

$$x_{ijk} = \mu + h_i + y_{ij} + e_{ijk}$$

where μ denotes the general mean

h_i denotes the effect common to the i th herd

y_{ij} denotes the effect common to the j th year
in the i th herd

e_{ijk} denotes the effect of permanent influences of the cow plus temporary environmental circumstances and random errors in evaluating the record.

The average of all available records was used in this study; thus, the observation for a given dam was:

$$\bar{x}_{ik} = \mu + h_i + \frac{\sum_j n_{ij} y_{ij}}{n_{i.k}} + \text{error}.$$

\bar{x}_{ik} is the average of all records of the k th dam in the

ith herd, and a dot in the notation indicates summation over that subscript. For example, $n_{i.k}$ is the total number of records made by the kth cow in the ith herd for all years (j) the cow was in the ith herd. In the right member of the equation the j or year notation has no meaning except in the term involving the year effect. For simplicity it is eliminated from all other terms. In the error for \bar{X}_{ik} , temporary environmental circumstances and the random error in evaluating the record would be divided by $n_{i.k}$, but the permanent influence of the cow would not be divided. In terms of components the expected or average values of the uncorrected total sum of squares, uncorrected sum of squares among herds, and the correction term are:

(a) Total

$$E \left\{ \sum_{ik} \bar{X}_{ik}^2 \right\} = n_{..} \mu^2 + n_{..} \sigma_h^2 + \sum_{ik} \left(\frac{\sum_j n_{ijk}^2 \sigma_y^2}{n_{i.k}} \right) + n_{..} \sigma_e^2$$

(b) Among herds

$$E \left\{ \frac{\sum_i \bar{X}_{i.}^2}{n_{i.}} \right\} = \sum_i \left\{ n_{i.} \mu^2 + n_{i.} \sigma_h^2 + \frac{\sum_j \left(\sum_k \frac{n_{ijk}^2}{n_{i.k}} \right) \sigma_y^2 + n_{i.} \sigma_e^2}{n_{i.}} \right\}$$

(c) Correction term

$$E \frac{\bar{X}_{..}^2}{n..} = n..\mu^2 + \sum_i n_{i.} \sigma_h^2 + \frac{\sum_{ij} \left(\sum_k \frac{n_{ijk}}{n_{i.k}} \right)^2}{n..} \sigma_y^2 + \sigma_e^2.$$

From these three basic equations the composition of each calculated sum of squares can be expressed in terms of the components of variance indicated in the original model. The error component as defined by the above model represents the intra-herd variance of the dams with an average of $n_{i.k}$ records, where each record has been corrected for yearly herd environmental effects.

In most dairy data a given dam may have several daughters with production records. Some additional information is furnished by the added daughters, but the information is not in direct proportion to the number of daughters. Inaccuracies in sampling the genes of the dam are reduced, but the failure of the dam's phenotype to reflect accurately the genes she possesses is not influenced. In these data the number of daughters per dam varied from one to six with a mean of 1.4. The dam's average was repeated for each of her daughters in computing the daughter-dam covariance. While including all daughters furnishes added information, the number

of degrees of freedom for estimating the intra-herd covariance is somewhat less than the total number of daughter-dam pairs minus the number of herds, although more than the total number of dams minus the number of herds.

Determining the intra-herd covariance between daughter and dam, free from the influence of years within herd, becomes somewhat involved in the actual computation, although it is straightforward in principle. Most of this tediousness results from using the average of all available records for the individual phenotype, and from the variation in the number of records which make up each cow's average.

Where g_{ik} is the average effect of the genes on fat production included as a portion of the error in the model on page 71, the equation for expressing the average phenotypic value of the dam is:

$$\bar{X}_{ik} = \mu + h_i + \sum_j \frac{n_{ijk} y_{1j}}{n_{ik}} + g_{ik} + \text{error}.$$

Likewise the equation for the daughters is:

$$Y_{ik_a} = \mu + h_i + \sum_j \frac{n_{ijk_a} y_{jk}}{n_{i.k_a}} + \frac{g_{ik_a}}{2} + \text{error}.$$

The subscript k_a indicates the a th daughter of the k th dam. Values for a go from one to six, and when summation over k is indicated for the dams the value for the k th dam is included for each of her daughters.

In terms of the components indicated in the preceding models the expected values of the uncorrected total cross products, the uncorrected cross products among herds, and the correction term are:

(a) Total

$$E \left\{ \sum_{ika} (\bar{X}_{ik}) (\bar{Y}_{ika}) \right\} = \sum_{ika} (n_{ika} \mu^2 + n_{ika} \sigma_h^2 + \frac{\sum_j (n_{ijk}) (n_{ijka}) \sigma_y^2 + \sigma_g^2}{n_{i.k} n_{i.ka}})$$

(b) Among herds

$$E \left\{ \frac{\sum_i \left(\sum_k \bar{X}_{ik} \right) \left(\sum_{ka} \bar{Y}_{ika} \right)}{n_i} \right\} = \sum_i (n_{i.} \mu^2 + n_{i.} \sigma_h^2 + \frac{\sum_j \left(\sum_k \frac{n_{ijk}}{n_{i.k}} \right) \left(\sum_{ka} \frac{n_{ijka}}{n_{i.ka}} \right) \sigma_y^2}{n_{i.}} + \frac{\sum_k n_{ik}^2 \frac{\sigma_g^2}{2}}{n_{i.}})$$

(c) Correction term

$$E \left\{ \frac{(\bar{X}_{..}) (\bar{Y}_{..})}{n_{..}} \right\} = n_{..} \mu^2 + \frac{\sum_i n_{i.}^2 \sigma_h^2}{n_{..}} + \frac{\sum_{ij} \left(\sum_k \frac{n_{ijk}}{n_{i.k}} \right) \left(\sum_{ka} \frac{n_{ijka}}{n_{i.ka}} \right) \sigma_y^2}{n_{..}} + \frac{\sum_{ik} n_{ik}^2 \frac{\sigma_g^2}{2}}{n_{..}}$$

From these equations the composition of the total, among herd, and within herd cross products can be expressed in terms of the components of covariance indicated by the models on page 74. The intra-herd covariance, when freed of the environmental contribu-

Table 11. Analysis of Variance for Dams
Where the Production Was Available

Source	d/f	Mean square	Expected composition
Total	3362	9,067	E + .571 Y + .995 H
Among herds	289	51,253	E + 3.605 Y + 11.572 H
Within herd	3073	5,100	E + .286 Y
Components			
H = 3840 E = 4952 Y = 515			

tion, is an estimate of half the genetic variance among the dams.

Tables 11 and 12 give the analyses for estimating the intra-herd variance of the dams and the intra-herd covariance between daughter and dam.

The estimate of Y from the analysis in Table 7 was substituted in the above mean squares and covariances. In all but three of the original 293 herds there were daughters with at least one complete lactation whose dams had also completed one lactation during

Table 12. Analysis of Covariance
for 4,764 Daughter-dam Pairs

Source	d/f	Covariance	Expected composition of covariance
Total	4762.2	4,721	$G/2 + .196 Y + .995 H$
Among herds	481.4	40,451	$G/2 + 1.847 Y + 9.839 H$
Within herds	4280.8	703	$G/2 + .010 Y$
Components			
	H = 3944	$G/2 = 698$	Y = 515

the period 1943 to 1947. Because the cows involved as either daughters or dams made up a large fraction of the total cows involved in the analysis of Table 7, the value of Y in this sample would probably not differ much from the estimate of Y assumed.

For the analysis of covariance the degrees of

freedom listed in Table 12 are the expected number of $G/2$ covariance components in the respective cross products. Where each dam has only one daughter in the analysis, n_{ik} in the last term of the equation for the cross product among herds is always one. In those cases the expected number of $G/2$ components in the total cross product is the number of dams minus one, in the cross product among herds it is the number of herds minus one, and in the cross product within herds the number of dams minus the number of herds. However, when several daughters of the same dam are included the above relations do not hold. The value of n_{ik} will vary from one to six in the present data. The expected number of $G/2$ components in the cross products among herds will be greater than the number of herds, then the expected number of $G/2$ components in the intra-herd cross products is necessarily less than the total number of daughters minus the number of herds. Disregarding this situation biases the estimate of the intra-herd covariance toward smallness. Where the intra-sire regression is used this is generally not so important, although there are some instances where a dam will have more than one daughter by the same sire. In other farm

animals, such as swine where full sisters are more abundant, this bias may be appreciable even on an intra-sire basis.

From the results indicated in Tables 11 and 12 the total regression of daughter on dam was .521, ignoring the influence of common herd environment. On an intra-herd basis, where the influence of year-to-year changes in herd environment was considered, the regression of daughter on dam was .141. The large difference between .521 and .141 emphasizes the importance of discounting the effect of herd differences before using a daughter-dam regression or correlation to estimate heritability. Ignoring the variance due to years, the estimate of the intra-herd regression is .138. By using the average of all records for daughters and dams, most of the confusing influence of yearly environmental changes on the estimate of the daughter-dam regression is averaged out. This is especially true with regard to the intra-herd covariance.

The regressions given above are based on averages of n records for the dams; consequently, they are larger than the expected regression calculated by using a

single random record of the dam. Lush and Straus (1942) have given a formula developed by Cochran to adjust regressions calculated on an average of \bar{n} records of the dam to the regression on a single record basis. Where b' is the regression calculated using averages, b is the regression of on the basis of a single record, \bar{n} is the mean number of records for the dams and v is the repeatability of lactations of the same cow, we have the following relation

$$b = b' \left\{ \frac{1 + (\bar{n}-1)v}{\bar{n}} + \frac{\sigma^2}{\bar{n}^3} (1-v) \right\}$$

If all dams had exactly \bar{n} records, it can be readily seen that the first term in the parenthesis would be sufficient. When \bar{n} varies for the different dams, as in the present instance, the second term is also necessary.

After the intra-herd regressions are adjusted to a single-record basis and multiplied by two, the estimates of the heritability of differences in single lactation records appropriately adjusted for year effects is .212. The estimate ignoring the influence of years is .201. The 95 percent confidence limits for the estimate of heritability ignoring the influence of years are .150

to .253. These seem wide in view of the amount of data used to estimate the regression; however, doubling the regression coefficient to estimate heritability also doubles the sampling errors.

No correction of the estimates of heritability could be made for the possible influence of inbreeding, since reliable information as to the degree of inbreeding among American Jerseys was not available. If the results of the studies of the amount of inbreeding in other breeds of American dairy cattle (Lush 1946 p. 332) are indicative of the situation in the Jersey breed, the correction for the effect of inbreeding is so small that it could be neglected with little difference in the result. Where the mating system is based on consanguinity and each herd is considered as a breeding group, inbreeding would have reduced the genetic variance of the dam from that expected under random mating. The covariance between daughter and dam would also be reduced by the same fraction. For the regression of daughter on dam the numerator of the coefficient would be reduced a bit more proportionately than the denominator. However, when the degree of inbreeding is no greater than indicated by

Lush (1946) the regression, and thus the heritability estimate would be influenced very little.

E. The Index

For traits, such as fat production, which may be expressed several times during an animal's lifetime, the phenotypes of the several individuals in the herd will be known with different degrees of accuracy depending largely upon the number of records for each individual. An index involving repeatable traits must take into account this variation in the accuracy with which the individual phenotypes are known, in addition to the variation in the amount of information furnished by the progeny. How would the attention given each item in the index shift if a cow's phenotype were based on an average of many records rather than on only one record? Similarly, when the number of daughters varies how does the emphasis on the progeny shift? The path coefficient analysis (Wright 1934) seemed most promising to obtain a general view of how the index would change as these items varied.

The path coefficient scheme considered in the present study is given in Figure 2. This diagram is intended to depict the situation existing among individuals kept in the same herd. Likewise, the index to be derived is intended for selection among animals in the same herd. The diversity of managerial regimes from herd to herd and the paucity of information concerning the portion of the inter-herd differences which are genetic make it difficult to develop an accurate method to discriminate between the breeding values of two individuals in different herds. Most choices of females involve intra-herd selection, but presumably most choices of one sire instead of another involve inter-herd selection.

Definition of Symbols

G : breeding value for a given character

U : combined effect of environment, dominance, and epistasis

\bar{X} : cow's phenotypic average¹

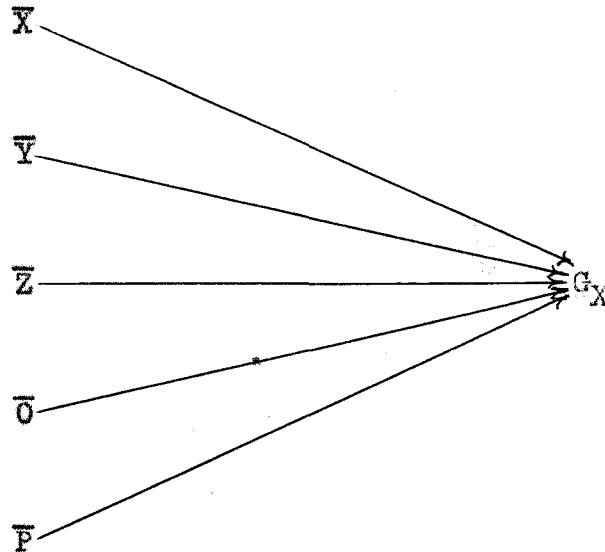
¹Individual phenotypes are expressed as a difference from the herd average.

- \bar{Y} : daughter's phenotypic average¹
 \bar{Z} : dam's phenotypic average
 \bar{O} : maternal half sister's phenotypic average
 \bar{P} : paternal half sister's phenotypic average
 ab : Mendelian segregation $1/2 \sqrt{\frac{1+f'}{1+f}}$
 c : genic correlation (relationship) between mates of
 a sire
 n : number of observations per individual (n_p), (n_x)
 etc. or
 number of individuals in a related group ($n_{\bar{p}}$), ($n_{\bar{y}}$),
 ($n_{\bar{o}}$).
 m : genic correlation between a sire and his mates; m'
 is the genic correlation between a sire and his
 mates for the preceding generation
 t : phenotypic correlation between members of a related
 group
 v : repeatability or correlation between different obser-
 vations of the same individual
 Other lower case letters indicate path coefficients

¹The group phenotypes may be considered either as an average or a sum when the correlations and standard partial regression coefficients are to be determined. If a different expression is more practical to use in the actual calculation of the index, this can be accommodated when the partial regression coefficients are determined.

Figure 2. Path Coefficient Diagram Showing the Biometric Relations Between a Cow's Phenotype (\bar{X}), Her Breeding Value (G_X), and the Phenotypes and Breeding Values of Her Close Relatives.

The appropriate prediction diagram, using the information given in Figure 2 is:



The prediction equation or index would then be

$$I = b_x \bar{X} + b_y \bar{Y} + b_z \bar{Z} + b_o \bar{O} + b_p \bar{P}$$

where the b 's are regression coefficients chosen to maximize $r_{G_X I}$ and the independent variables are as defined on pages 83 and 84.

Use of the correlation approach for devising an index requires expressing or evaluating the correlations between the breeding value (G_X) of the individual being indexed, and each of the phenotypic observations used as independent variables. All possible correlations

between those phenotypic observations are needed also. These correlations were derived in general terms using Wright's method of path coefficients. The two general uses of the path coefficient method are to analyze a correlation into its various parts and to express the degree to which the variance of the dependent variable is determined by each of the causal factors. A concise statement of the application of path coefficients for these two purposes is given in the article by Wright (1921, p. 114-115) and restated in detail by him later (Wright 1934).

For illustration a portion of Figure 2 is reproduced in Figure 3 to show the biometric relation between the phenotype of the dam and that of her daughters. A simple case is illustrated in (a) of Figure 3 where the dam has one record and she in turn has only one daughter with a single record.

For the situation in (a) of Figure 3:

$$r_{XY} = g^2_{ab} (1 + m)$$

A more general situation is illustrated in (b) of Figure 3.

The dam has n_x records and she has n_y daughters each with n_y records. In this more general case

$$\begin{aligned} r_{XY} &= g^2 ab n_x \times n_y \sqrt{n_y} \sqrt{n_y} (1+m) \\ &= g^2 ab \sqrt{\frac{n_x}{1+(n_x-1)v}} \sqrt{\frac{n_y}{1+(n_y-1)v}} \sqrt{\frac{n_y}{1+(n_y-1)t_{yy}}} (1+m). \end{aligned}$$

A list of the needed correlations between the items in Figure 2 are given below.¹

$$\begin{aligned} r_{GX} &= g \sqrt{\frac{n_x}{1+(n_x-1)v}} \\ r_{GY} &= gab \sqrt{\frac{n_y}{1+(n_y-1)v}} \sqrt{\frac{n_y}{1+(n_y-1)t_{yy}}} (1+m) \\ r_{GZ} &= gab \sqrt{\frac{n_z}{1+(n_z-1)v}} (1+m') \end{aligned}$$

¹If a combination of consanguine and phenotypic assortive mating is being practiced within a herd, neglect of some minor terms may not permit some of the correlations to be considered exact. However, these terms will be so small in most circumstances that their influence on the actual values of the correlations generally can be neglected. In both Figure 2 and in the correlations given here the relationship between different sires mated to a particular cow is assumed to be zero. Likewise, the relationship between the sires of one generation and the sires and dams of the preceding generation is assumed to be zero.

$$r_{G_{x\bar{O}}} = g a^2 b^2 \sqrt{\frac{n_o}{1+(n_o-1)v}} \sqrt{\frac{n_{\bar{o}}}{1+(n_{\bar{o}}-1)t_{oo}}} (1+2m')$$

$$r_{G_{x\bar{P}}} = g a^2 b^2 \sqrt{\frac{n_p}{1+(n_p-1)v}} \sqrt{\frac{n_{\bar{p}}}{1+(n_{\bar{p}}-1)t_{pp}}} (1+2m'+c)$$

Phenotypic Correlations

$$r_{\overline{XY}} = g^2 a b \sqrt{\frac{n_x}{1+(n_x-1)v}} \sqrt{\frac{n_y}{1+(n_y-1)v}} \sqrt{\frac{n_y}{1+(n_y-1)t_{yy}}} (1+m)$$

$$r_{\overline{XZ}} = g^2 a b \sqrt{\frac{n_x}{1+(n_x-1)v}} \sqrt{\frac{n_z}{1+(n_z-1)v}} (1+m')$$

$$r_{\overline{XO}} = g^2 a^2 b^2 \sqrt{\frac{n_x}{1+(n_x-1)v}} \sqrt{\frac{n_o}{1+(n_o-1)v}} \sqrt{\frac{n_{\bar{o}}}{1+(n_{\bar{o}}-1)t_{oo}}} (1+2m')$$

$$r_{\overline{XP}} = g^2 a^2 b^2 \sqrt{\frac{n_x}{1+(n_x-1)v}} \sqrt{\frac{n_p}{1+(n_p-1)v}} \sqrt{\frac{n_{\bar{p}}}{1+(n_{\bar{p}}-1)t_{pp}}} (1+2m'+c)$$

$$r_{\overline{YZ}} = g^2 a^2 b^2 \sqrt{\frac{n_y}{1+(n_y-1)v}} \sqrt{\frac{n_{\bar{y}}}{1+(n_{\bar{y}}-1)t_{yy}}} \sqrt{\frac{n_z}{1+(n_z-1)v}} (1+m+m')$$

$$r_{\overline{YO}} = g^2 a^3 b^3 \sqrt{\frac{n_y}{1+(n_y-1)v}} \sqrt{\frac{n_{\bar{y}}}{1+(n_{\bar{y}}-1)t_{yy}}} \sqrt{\frac{n_o}{1+(n_o-1)v}} \sqrt{\frac{n_{\bar{o}}}{1+(n_{\bar{o}}-1)t_{oo}}} (1+m+2m')$$

$$r_{\overline{YP}} = g^2 a^3 b^3 \sqrt{\frac{n_y}{1+(n_y-1)v}} \sqrt{\frac{n_{\bar{y}}}{1+(n_{\bar{y}}-1)t_{yy}}} \sqrt{\frac{n_p}{1+(n_p-1)v}} \sqrt{\frac{n_{\bar{p}}}{1+(n_{\bar{p}}-1)t_{pp}}} (1+m+2m'+c)$$

$$r_{\overline{ZO}} = g^2 a b \sqrt{\frac{n_z}{1+(n_z-1)v}} \sqrt{\frac{n_o}{1+(n_o-1)v}} \sqrt{\frac{n_{\bar{o}}}{1+(n_{\bar{o}}-1)t_{oo}}} (1+m')$$

$$r_{\overline{ZP}} = g^2 a b \sqrt{\frac{n_z}{1+(n_z-1)v}} \sqrt{\frac{n_p}{1+(n_p-1)v}} \sqrt{\frac{n_{\bar{p}}}{1+(n_{\bar{p}}-1)t_{pp}}} (m'+c)$$

$$r_{\overline{OP}} = g^2 a^2 b^2 \sqrt{\frac{n_o}{1+(n_o-1)v}} \sqrt{\frac{n_{\bar{o}}}{1+(n_{\bar{o}}-1)t_{oo}}} \sqrt{\frac{n_p}{1+(n_p-1)v}} \sqrt{\frac{n_{\bar{p}}}{1+(n_{\bar{p}}-1)t_{pp}}} (m'+c)$$

Values for the b 's in the index illustrated on page 86 are obtained by first solving the system of normal equations in standard measure indicated below.

$$\begin{aligned}
 B_x + r_{XY}B_y + r_{XZ}B_z + r_{XO}B_o + r_{XP}B_p &= r_{G_xX} \\
 r_{YX}B_x + B_y + r_{YZ}B_z + r_{YO}B_o + r_{YP}B_p &= r_{G_xY} \\
 r_{ZX}B_x + r_{ZY}B_y + B_z + r_{ZO}B_o + r_{ZP}B_p &= r_{G_xZ} \\
 r_{OX}B_x + r_{OY}B_y + r_{OZ}B_z + B_o + r_{OP}B_p &= r_{G_xO} \\
 r_{PX}B_x + r_{PY}B_y + r_{PZ}B_z + r_{PO}B_o + B_p &= r_{G_xP}
 \end{aligned}$$

The B 's from the solution of the above equations give the b 's, or partial regression coefficients, by the relation $b_x = B_x \frac{\sigma_G}{\sigma_X}$.

A general solution for the B 's was not easily obtainable. The solution in any case would have been so complex that the utility of the expressions for the B 's would be practically nil. Even in a simple case when only information on the cow (\bar{X}) and her daughter (\bar{Y}) is to be considered in the index, the expressions for the B 's are clumsy. In the case shown in Figure 3 (b), where the cow's phenotype and the phenotypic average of her daughters are used as an index and $m = 0$, but the expression is kept general for the number of

records per individual and number of daughters

$$\beta_x = \frac{8 \sqrt{\frac{n_x}{1+(n_x-1)v}} \sqrt{1-g^2 a^2 b^2} \left\{ \left(\frac{n_y}{1+(n_y-1)v} \right) \left(\frac{n_{\bar{y}}}{1+(n_{\bar{y}}-1)t_{yy}} \right) \right\}}{1-g^4 a^2 b^2 \left(\frac{n_x}{1+(n_x-1)v} \right) \left(\frac{n_y}{1+(n_y-1)v} \right) \left(\frac{n_{\bar{y}}}{1+(n_{\bar{y}}-1)t_{yy}} \right)}$$

and

$$\beta_y = \frac{8ab \sqrt{\frac{n_y}{1+(n_y-1)v}} \sqrt{\frac{n_{\bar{y}}}{1+(n_{\bar{y}}-1)t_{yy}}} \left\{ 1-g^2 \left(\frac{n_x}{1+(n_x-1)v} \right) \right\}}{1-g^4 a^2 b^2 \left(\frac{n_x}{1+(n_x-1)v} \right) \left(\frac{n_y}{1+(n_y-1)v} \right) \left(\frac{n_{\bar{y}}}{1+(n_{\bar{y}}-1)t_{yy}} \right)}$$

It seemed more logical to evaluate the correlations involved in a given situation and then solve the system iteratively for the β 's. After solving several special cases the consequence of variation in number of records for each individual and in number of individuals in a progeny average could be viewed.

For simplicity in evaluating the correlations, the path coefficient from breeding value of parent to breeding value of offspring is considered to be one-half.

Wright (1921) has shown that ab equals $\frac{1}{2} \sqrt{\frac{1+f'}{1+f}}$ where f' and f refer to the inbreeding coefficients of the parents

and offspring respectively. Unless the population is being inbred rapidly or is being outcrossed widely, f and f' will be almost exactly the same and the value of ab is one-half.

The estimates of repeatability, paternal half sister correlation, maternal sister correlation and heritability used to evaluate these correlations were those computed where the yearly influences were ignored. Ignoring those makes the findings more representative of the situation under practical conditions where much of the year-to-year variation is left in the individual phenotypes. The estimates are repeatability, .412; paternal half sister correlation, .120; maternal half sister correlation, .086; and heritability, .201. The genic correlation between sire and dam (m) was assumed to be zero. It cannot be far from zero if most of the sires used in a herd are purchased from different breeders. The genic correlation (relationship) between dams or the mates of a sire (c) is assumed to be .1. No estimate of c was made from the dams' pedigrees in the present study. However, when most of the dams are born in the herd where they produce their offspring, c

probably will have a real value. Lush (1942) has indicated that the average relationship between mates of a sire would be 8 to 12 percent if one-third are half sisters, one-third are three-quarter sisters unrelated to the half sisters, and there are a few sets of full sisters or daughter-dam pairs. This would be fairly close to the general situation in the many herds where only one sire is used at a time, and he is kept in service for about two years, no line breeding is practiced, and the females are all raised from heifers born in the herd. Relationship would be less in larger herds or where females continue to be purchased, but it would be more in the few herds which practice a considerable amount of line breeding.

The average of the deviation of all records from the herd average is considered the phenotype of an individual in the diagram of Figure 2. In the progeny averages, each individual is assumed to have the same number of records, but the average number can range from 1 to n in specific cases covered by the diagram. If the individual phenotypic averages making up the group average are not equally variable (number of

observations for each individual not the same), the paths from the individual phenotypes to the group average do not all have the same value. When several records of the same cow were averaged, the assumption that each record was equally variable and had proportionately the same influence on the average seems to be reasonable (Berry 1945 p. 356). However, the above assumption is not tenable for the progeny average unless each individual has the same number of records. Where the individuals have different numbers of records, the situation could be remedied by using only a single unselected record for each individual (e.g. the first record). This would be discarding some of the available information; thus, it is not an efficient solution to the problem.

The variance of averages of n records is $\frac{1+(n-1)v}{n}$ times the variance of single records. An average of \underline{n} records of a cow more accurately evaluates the true productiveness of the cow than does a single record. Thus, if cows with one, two, or more records are being combined for a progeny average or sum, the average of each cow should be weighted by the reliability with which that average indicates the cow's true average. This can be

considered in terms of the information contained in each average which Fisher (1947) has designated $\frac{1}{\sigma^2}$. Each cow's average deviation from the herd mean should be multiplied by $\frac{n}{1+(n-1)v}$ to be properly weighted.

This weighting can be accomplished simply by using for each cow's phenotype her estimated real producing ability instead of her actual average. Lush (1945) has shown that a cow's real producing ability as a deviation from the herd average is equal to

$$\frac{nv}{1+(n-1)v} \text{ times (her average minus the herd average).}$$

When all individual phenotypes are expressed as estimated real producing abilities, the number of possible combination of the independent variables which need to be considered in the index is greatly reduced. The weighting takes care of different numbers of observations for each individual, and only the problem of different numbers of individuals per sire or dam persists.

The choice of the proper herd average is a complex problem. For cattle the yearly herd average will be subject to much sampling variation if the number of cows

per herd is small. For this reason alone, the yearly herd average will not be a wholly reliable estimate of the true herd mean. Furthermore, year-to-year changes in the herd average are a function both of changes in herd environment and of changes in the cow composition. Expressing each record as a deviation from the yearly herd average introduces an uncertain amount of true cow difference into the comparisons or contrasts between records of the same cow. This quantity will vary from situation to situation and would be more pronounced for herd averages made up of records from only a few cows.

Since the number of cows per herd is usually small, it is suggested that a moving average of the present year plus the previous three or four years be used. For each successive year the production for the earliest year would be replaced by the present year's production. This procedure will introduce more yearly differences into comparison among cows than if the yearly herd averages were used; however, the part played by cow differences in changing the herd average would be decreased. Since these data indicate C to be over five times as large as Y, the moving average is a reasonable, though

admittedly not a perfect solution.

1. Index including cow's own production.

The general correlations given on page 90 were evaluated and a number of solutions were made to determine the relative value of information on the individual's own performance and the performance of its relatives for predicting the individual's breeding value. Increase in information on one variable will alter the attention paid to other variables, depending on the correlations between them. To note the importance of different numbers of individuals per sire or dam, solutions were made where information on one of the variables shifted, e.g. the number of daughters, but the information on the other variables remained the same. Each variable in turn was treated in a similar manner. Solution of a number of specific situations gave the B 's which were converted to the b 's. Each of the partial regression coefficients was multiplied by the constant 2.7027 since this made the value for b_x equal to 1. The index for cows with production records then is:

$$I^1 = \overset{V}{X} + b_y \overset{V}{Y} + .42 \overset{V}{Z} + b_o \overset{V}{O} + b_p \overset{V}{P}$$

Values for b_y , b_o , and b_p are given in Table 13 for different numbers of individuals in each sum. The value of b_x has been adjusted to permit including the individual being indexed in the paternal half sister sum. This eliminates the necessity of computing the sum for paternal half sisters anew for each individual. Considerable complexity in the index is avoided by including full sisters simply as maternal and again as paternal half sisters.

The phenotypic values for the paternal half sisters, the maternal half sisters, and the daughters can be expressed as sums of each individual's real producing ability. The rapid increase in the standard deviation of the sum permits the partial regression coefficients to take on more nearly constant values over the range of variation in number of individuals in each group than would the regression coefficients if averages instead of sums were used for the paternal and maternal sisters.

¹A $\overset{V}{}$ indicates an individual's most probable producing ability. For $\overset{V}{Y}$, $\overset{V}{O}$, and $\overset{V}{P}$ the sum of the individual most probable producing abilities is involved.

Table 13. Values of b_y , b_o , and b_p
for Various Values of $n\bar{y}$, $n\bar{o}$, and $n\bar{p}$

$n\bar{y}$	b_y	$n\bar{p}^a$	b_p
1	.405	1	.148
2	.380	2	.236
3	.346	3	.203
4	.318	4	.179
5	.300	5	.159
		6	.143
		7	.130
		8	.120
		9	.110
		10	.103
$n\bar{o}$	$b\bar{o}$		
1	.151	11	.096
2	.138	12	.090
3	.126	13	.085
4	.117	14	.080
5	.110	15	.076
		--	--
		--	--
		--	--
		20	.060

^aThe individual being indexed is added in $n\bar{p}$; thus when $n\bar{p}$ is equal to one, \bar{X} and \bar{P} are the same value.

Few cows will have more than one or two daughters; thus using b_y equal to .4 and b_o equal to .14 as constants for all situations would introduce little error but much simplification in the actual use of the index. The regression coefficients for the paternal half sisters shift a bit more than b_o or b_y over the probable range of variation in n_p . In practice, some simplification might be introduced by using four or five values of b_p rather than some 15 or 20 as are shown in Table 13. A compromise on values of b_p for different numbers of paternal half sisters might be .15 for one (actually the cow being indexed), .22 for two or three, .16 for four to six, .12 for seven to ten, and .09 for over ten.

Values of r_{G_I} for the above index will shift as the amount of information included in the index varies. Where the cow has three records, her dam has three records, the cow has one daughter with one record, one maternal half sister with two records and four paternal half sisters with two records each, r_{G_I} is .656 as contrast to r_{G_X} of .576 where selection was solely on own phenotype. This would mean that progress would be about 1.12 times more rapid by using, in addition to

the cow's records, the information on her relatives. Where the cow has fewer records, the same information on the relatives is comparatively more important. Conversely, if the cow has more records the comparative value of the relatives is reduced.

If heritability were lower than .201, the relatives would assume more weight in determining the individual's index value. Assume heritability is .15, which is the lower five per cent fiducial limit of the estimate from these data. Then the index where the cow had one daughter, one maternal half sister and four paternal half sisters should be:

$$I = .86 \overset{V}{X} + .40 \overset{V}{Y} + .41 \overset{V}{Z} + .16 \overset{V}{O} + .15 \overset{V}{P}.$$

For the same situation but where heritability is assumed to be .25, which is the upper five per cent fiducial limit, the index should be:

$$I = 1.11 \overset{V}{X} + .40 \overset{V}{Y} + .42 \overset{V}{Z} + .13 \overset{V}{O} + .16 \overset{V}{P}.$$

With higher values of heritability the individual's own phenotype assumes a greater role in determining the index value. There is, however, not a marked shift for the

range of values for heritability considered here.

2. Index for heifers or young bulls.

Own production or daughters production is not available for heifers or young bulls. An animal must be indexed solely on the performance of its relatives. The diagram in Figure 2 will apply to this situation if we delete the information furnished by the individual's phenotype (\bar{X}) and its daughters (\bar{Y}). Since the individual's phenotype assumed the major role in determining the index for cows in milk, the values of the β 's for the dam, maternal and paternal half sisters in the solution for the index of cows in production differ from those obtained where only the above three items are included for a solution. Genetic progress realized from selecting the young stock by using an index developed separately for individuals without production records would be approximately 1.20 times that realized by using the weightings of the index for cows with production records. Although introducing a separate index for the younger animals complicates the practical application of the index, the additional gain from using the index seems to justify its introduction. This index for

young animals is:

$$I = .63 \overset{V}{Z} + b_o \overset{V}{O} + b_p \overset{V}{P}$$

where the values for b_o and b_p are determined by the number of maternal and paternal half sisters for each individual. Table 14 gives values for b_o and b_p for

Table 14. Values of b_o and b_p for Various Values of $n\bar{o}$ and $n\bar{p}$.

$n\bar{o}$	b_o	$n\bar{p}$	b_p	$n\bar{p}$	b_p
1	.236	1	.353	10	.170
2	.220	2	.316	11	.160
3	.197	3	.285	12	.152
4	.181	4	.260	13	.145
5	.170	5	.239	14	.138
		6	.220	15	.132
		7	.205	--	--
		8	.192	--	--
		9	.180	20	.108

different numbers of maternal and paternal half sisters. For practical use a single value for b_o could be used and some compromise to simplify use of values for b_p should be considered.

Efficiency of pedigree selection does not approach

that of individual selection for most situations, even if heritability is as low as .201. When available, the phenotype of the individual should assume a major role in determining an individual's index value. The precision with which the individual breeding values of young animals can be predicted is low in any case. Where the dam of a heifer has three records, one maternal half sister has one record and four paternal half sisters have one record each, rg_{XI} is only .361. This may be contrasted with rg_{XX} , which is .449 if the individual has one production record. Only when proof on the sire and dam is extensive does the pedigree information approach the value of a single record for predicting an individual's breeding value. Where own production is not available, rg_{XI} is .396 when the dam of the heifer has five records, two maternal half sisters have an average of two records each and four paternal half sisters have one record each.

The variance of an index can be derived by using the simple relations for the variance of a sum where the b 's are constants. The variance of the index for heifers can be expressed:

$$\sigma_I^2 = b_z^2 \sigma_v^2 + b_o^2 \sigma_v^2 + b_p^2 \sigma_v^2 + 2 \text{Cov } ZO + 2 \text{Cov } ZP + 2 \text{Cov } OP.$$

The variance of the index will vary with the quantity of information available concerning each individual. If heifers have dams with an average of two records, one maternal sister with one record, and four paternal sisters with one record each, the variance of their index values would be expected to near 1000. For cows in milk averaging two records each with dams averaging three records, and the same information from half sisters as in the preceding example, the index values would be expected to have a variance of approximately 3000.

Indexes for cows with production records will be more variable than those for young heifers. Selection should be made on the size of the index without regard to whether there is much or little information on that individual. The index will take care of that automatically, making the values conservative where the pertinent information is scanty (r_{GI} is low) and making them vary widely when r_{GI} is large. Both the highest and the lowest indexes generally will be for cows which already have records. If further freedom to cull existed after the

cows with the very lowest indexes were removed, some of the young animals would be removed next. Where the production of a cow is so poor that the question of culling her daughters arises, genetic progress would be maximized by culling the cow herself. Her daughter is likely to be nearer the herd average in breeding value.

In regions where the cost of raising a heifer to milking age is more than her beef value, this economic circumstance may dictate culling unpromising heifers a bit more strictly than one culls the cows which have the same index level. Since culling of cows without own production is less accurate than culling after own production is available, the rate of genetic improvement may be slowed down by that but perhaps not enough to make it worthwhile to incur the extra cost of raising the questionable heifers to determine more accurately which of them should be culled. The cost of raising already has been incurred for the cows, and under some circumstances the net profit might be greater by continuing to milk some poor cows through the prime of their life, while culling instead some young heifers whose indexes actually were a bit higher than those for these cows already in production.

V. DISCUSSION

Accurate estimates of the basic items are essential for constructing an efficient index. For the present study these fundamental items for fat production are repeatability, heritability, maternal half sister correlation, and paternal half sister correlation. Intra-herd repeatability of records of the same cow seems to be close to .4 where a comparatively unselected population of cows are considered. The estimate of .41 from the present data agrees closely with most of the estimates from other data given in Table 1. Although intra-herd year-to-year variations in general environment caused only a small fraction of the total intra-herd variability, repeatability could be increased nearly ten percent if records could be corrected for this variation.

With slightly over 4000 degrees of freedom to estimate the sire component, a reliable estimate of the paternal half sister correlation was obtained. The maternal half sister correlation was not estimated within such narrow limits, as there were only 1543 degrees of

freedom to estimate the variance from differences between maternal half sisters. However, after accounting for the fact that some of the intra-herd comparisons between dams involved paternal half sisters, there seems to be no consistent bias in the estimate of the maternal half sister correlation.

Compared to the other statistics the estimate of heritability has a somewhat greater sampling error because the observed regression had to be multiplied by two. However, the relative weight given the individual's performance and the information on its relatives does not vary much if heritability is between .15 and .25, and this is the 95 percent confidence interval for the present estimate of .201. Hence, over this range for heritability the ratios in the index calculated for heritability at .201 are not seriously in error. The value for heritability obtained in the present study is about in the middle of the range of estimates presented in Table 2.

Altogether 4971 daughter-dam pairs were available where both daughter and dam had at least one record. For 207 of these pairs the daughter and the dam had

records in different herds. These were not used to estimate heritability. The data were on IBM cards so that the herd in which the daughter made her records was the basis for classifying the daughter-dam pairs by herds. If the pairs where the daughter and dam made their records in different herds had not been eliminated, the influence of herd differences would not have been completely removed from the variance of the dams. With a herd component as large as that found in these data, and with the present proportion of the total daughter-dam pairs where the daughter and dam made their records in different herds, failure to eliminate these 207 pairs probably would have lowered the heritability estimate by .01 or .02.

When bred heifers are sold into other herds, they and the heifer calves they are sometimes carrying may make their records in those herds. However, these herds would be different from the herd where the other daughters and mates of a sire made their records. Although a sire actually may have been used in only one herd, the sale of heifers bred to him would permit the resulting situation to be comparable to that where the sire had been used in several herds. The possibility of such circumstances further points to the necessity of making the

intra-sire regression also an intra-herd regression to be certain herd differences are not inflating the estimates obtained from the data.

Year-to-year environmental variation within herd accounted for only a small part of the total variance. Where the variance due to years is no larger than in the present data, its influence on the estimates of paternal half sister correlation, maternal half sister correlation, and heritability is small. For the regression of daughter on dam, using averages of all records for daughter and dam reduced the amount of year component in the intra-herd covariance to an almost negligible quantity.

In like manner, using the average of several records, rather than single records, in computing the index corrects for most of the confusing influence of year-to-year environmental variation. If a long time trend in the herd average exists, nevertheless, this cannot be corrected so simply. Unless some adjustment is made for any important changes in the herd average there may be considerable bias in the comparisons between young and old cows. The problem cannot be

solved in a straightforward manner until the trend in the herd average can be separated into hereditary and environmental portions.

The question arises as to how closely the results in a specific situation will follow those predicted for the population in general. For the present case one might wonder if the estimates of the basic parameters are actually applicable to a particular herd. There seems to be no clear-cut means for discriminating between the estimates from different herds. Where the basic units are so small, sampling errors alone may permit impossibly large or small answers. Accuracy of prediction of the results in a single herd for a short period, such as the next five years, naturally is very low but the pluses in some herds and the minuses in other herds should cancel each other, so that the prediction of the average outcome in a whole breed or other population of many herds may reach a high level of accuracy. Assuming the breeding system does not change extremely and suddenly, results from a generation of selection in a particular herd may be lower or higher than expected; but the predicted and actual results

should approach one another more closely if the period is extended over several generations.

It may be suggested by some that inaccuracy would be introduced by using an index based on fat alone if fat test and milk yield are inherited as separate characters. Lush et al (1941), however, have shown that where increased total fat is desired the error in assuming a single index for fat instead of using indexes for milk and fat percent separately is negligible.

A selection or prediction scheme involving only linear relationships is most effective if a major portion of the hereditary variance is due to additive or average effects of genes. Epistasis, dominance, or perhaps overdominance in fat production may contribute to what has been termed "specific combining ability". Many accounts of "nicking" are published in the literature, but usually the information, in addition to being selected, is meager in scope. With repeatability of fat production approximately .40 and heritability approaching .20, one wonders what portion of the variance between cows is accounted for by dominance or epistasis. Seath and Lush (1940) did not find evidence that dominance or

epistasis were important contributors to the observed variability in a population that is essentially random bred. However, where enough inbreeding is practiced to make families distinct, the effects of dominance and epistasis would be magnified. Probably most of the discrepancy between a cow's breeding value and her real producing ability is attributable to permanent environmental influences which affect the cow's production through many lactations (Berry 1945).

The index developed in the present study considers only one item of economic importance. Presumably that is the most important item. The appearance of the cow, her physical conformation or type also has economic importance. Assigning an appropriate figure for the economic importance of type is an elusive problem. In some highly advertised breeding herds where many of the customers may have ambitions to enter show-ring competition, type may be an item that contributes much to the total income. In striking contrast, type of itself may be of almost no importance in many grade herds. The attention given type in an efficient breeding program in addition to its relative economic importance, would be

determined by the heritability of type and by the phenotypic and genetic correlations between type and the other traits considered.

Harvey (1949) found the phenotypic correlation between type and production to be .12 for these cows which were included in the present study and also had an official type rating. The genetic correlation was estimated to be between .15 and .20 and the heritability of type was estimated as .14. All of these statistics, in addition to the relative economic value of type as compared to production, are required to incorporate type into an efficient index.

Other characteristics may be included in an index effectively if the same basic constants needed for the inclusion of type can be estimated for them. Increased breeding or reproductive efficiency may be important in increasing the economic return from a dairy herd, though it may or may not appreciably affect the average breeding value of the herd when measured only in terms of the lactation fat yield. Delayed breeding and the accompanying prolonged interval between calving greatly reduces the average herd production per calendar year

although these things might even raise the average production per 305-day lactation. Some culling is practiced automatically against delayed breeding, but its intensity has not been accurately determined. The genetic differences in reproductive efficiency among cows might play such a minor role that selection would be only slightly effective. At any rate, some of the problems concerning this topic still need clarifying.

- Including several traits automatically lessens the intensity of selection that can be practiced for any one trait. Nevertheless, if the conception of the aggregate breeding value is accurate, simultaneous selection for the several traits will permit the breeder to make the greatest economic gain from his breeding program.

Although the selection index provides a powerful tool for objective selection, it cannot hope to cover all situations in any simple manner. Including many traits and information on several relatives increases the complexity of the index. However, with dairy cattle where the individual unit represents so much value and the generation interval is so long, more actual time spent in calculating indexes seems to be warranted than for most species of farm livestock.

Use of an index will do much toward making selection objective. Such a procedure is needed especially in dairy cattle where there is much opportunity for sentiment to replace sound judgement. The index developed in the present study is only a beginning and information on other traits should be included. However, an index which would cover accurately all situations and include all pertinent traits would surely be unworkable because of its complexity. Individual judgement probably never can be replaced completely by any index or other mathematical concept for selection, but it can be relegated to a more minor role than it plays now.

VI. SUMMARY

The purpose of the present study was to develop a selection index for fat production where the fat yields of the cow and her relatives are used. Construction of the index called for reliable estimates of four fundamental statistics. These four were repeatability or intraclass correlation between records of the same cow, correlation between paternal half sisters, correlation between maternal half sisters, and heritability.

Data to estimate these statistics were taken from 293 Jersey herds on HIR test during the years 1943 to 1947. A total of 23,330 lactations records from 12,405 cows was included in the study. The average 305-day 2x M.E. fat production was 429 pounds. The 293 herds were scattered among 42 states.

The analysis of variance, including the separation of the total variance into its additive components, was the principal analytical method. Correlations between the items included in the index were derived using Wright's method of path coefficients. The weights to be

given the fat yields of the individual and its relatives were determined by multiple correlation.

Repeatability was estimated to be .412 with ten percent fiducial limits of .399 and .422. When the variance caused by year-to-year environmental variation within herd was removed, the estimate of repeatability rose to .459. Presumably this value of repeatability could not be realized in practice unless the variation attributable to environmentally caused fluctuations of the herd average from year-to-year could be removed from the records which actually are used.

When intra-herd yearly variation was accounted for, the correlation between paternal half sisters was .123. Ignoring the year component of variance the estimate was .120. There seemed to be a strong environmental contribution to these correlations although its basis is not clear. The estimate of the maternal half sister correlation was .090 when the influence of the year component was considered and when correction was made for some of the daughters of the different dams being paternal half sisters. Ignoring the influence of the year component made the estimate .086.

Heritability was estimated by doubling the intra-herd regression of daughter on dam as derived from 4,764 daughter-dam pairs from 290 herds. The 4764 daughters were from 3363 different dams. Heritability of differences in fat production on a single record basis was

Table 15. Values for the Several Components of Variance Computed from These Data

Component	Value	Component	Value
H	4060	HY	462
C	2730	S	741
E	3214	D	556
Y	515	P	2090
\bar{Y}	55	M	2282

estimated as .201. The five percent fiducial limits for this estimate are .150 and .253.

The total variance was apportioned into its several components in the various aspects of the study. Table 15 summarizes the values found for the components in these data. In the repeatability analysis the total variance was separated into H due to gross herd differences, C for permanent differences between cows within the herds, Y

for environmental and management changes which applied to all the cows in a herd in a given year but varied from year to year, and E for temporary differences other than those accounted for by Y which applied to one cow but not to all her herd mates and varied for the several lactations of the same cow. The component Y was broken down further into \bar{Y} for the average influence of any general year trend over all the 293 herds and HY for true herd x year interaction. To determine the correlation between paternal half sisters, variance due to permanent differences among cows was separated into S for differences among sires and P for differences among paternal half sisters. A second breakdown of the variance for permanent differences among cows was made to determine the correlation between maternal half sisters. In this analysis the variance associated with those differences was separated into D for variation among dams and M for permanent differences among maternal half sisters.

Indexes were constructed for cows with production records, and for young animals where own phenotype was not available. The relatives considered in the index

were the cow's dam, daughters, maternal half sisters and paternal half sisters. If heritability is near .20, which is the value found in these data, information on the individual should receive about 2.75 times as much attention as the same information on the dam. The relative weight to be given information on the other relatives varies with the number of individuals making up the sum and in some cases whether some of the relatives are present.

Progress to be expected by using the index for selections would be about 1.10 to 1.15 times faster than by making the selections on own performance alone. However, that ratio would depend on the number of records on the individual, on the number and kind of relatives, and on the amount of information about each.

VII. LITERATURE CITED

- Berry, J. C. 1945. Reliability of averages of different numbers of lactation records for comparing dairy cows. *Jour. Dairy Sci.* 28:355-366.
- Bonnier, G. 1935. Is the shape of the lactation curve genetically determined? *Hereditas.* 20:199-213.
- _____ 1946. The sire index. *Acta Agriculturae Suecana.* 1:321-334.
- Crump, S. Lee. 1946. The estimation of variance components in analysis of variance. *Biometrics Bulletin.* 2:7-11.
- Dickerson, G. E. 1940. Estimates of producing ability in dairy cattle. *Jour. Agr. Res.* 61:561-586.
- Eisenhart, C. 1939. The interpretation of certain regression methods and their use in biological and industrial research. *Ann. Math. Stat.* 10:162-186.
- Eldridge, Franklin E. and Salisbury, G. W. 1949. The relation of pedigree promise to performance of proved Holstein-Friesian bulls. *Jour. Dairy Sci.* 32:841-848.
- Fisher, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Trans Roy. Soc. Edinburgh.* 52:399-433.
- _____ 1936. The use of multiple measurements in taxonomic problems. *Ann. Eugenics.* 7:179-189.
- _____ 1946. Statistical methods for research workers. 10th ed. London, Oliver and Boyd.
- _____ 1947. The design of experiments. 4th ed. New York, Hafner Publishing Co., Inc.

- Gavin, W. 1913. Studies in milk secretion: on the accuracy of estimating a cow's milking capacity from her first lactation yield. Jour. Agr. Sci. 5:377-390.
- Gowen, J. W. 1920a. Studies in milk secretion. V. On the variations and correlations of milk secretion with age. Genetics 5:111-188.
- _____ 1920b. Studies in milk secretion. VI. On the variations and correlations of butterfat percentage with age in Jersey cattle. Genetics 5:249-324.
- Harris, G. M., Lush, J. L., and Shultz, E. N. 1934. Progress report on comparison of lactation and yearly records. Jour. Dairy Sci. 17:737-742.
- Harvey, W. R. 1949. Genetic variation and covariation in type and butterfat among Jersey cattle. Unpublished Ph.D. thesis, Ames, Iowa, Iowa State College Library.
- Hazel, L. N. 1943. The genetic basis for constructing selection indexes. Genetics 28:476-490.
- _____ and Lush, Jay L. 1943. The efficiency of three methods of selection. Jour. Heredity 33:393-399.
- _____ and Terrill, Clair E. 1946. The construction and use of a selection index for range Rambouillet lambs. (Abstract) Jour. Anim. Sci. 5:412.
- Henderson, C. R. 1949. Estimation of changes in herd environment. (Abstract) Jour. Dairy Sci. 32:706.
- Hetzer, H. O., Dickerson, G. E., and Zeller, J. H. 1944. Heritability of type in Poland China swine as evaluated by scoring. Jour. Anim. Sci. 3:390-398.
- Johansson, I. and Hansson, A. 1940. Causes of variation in milk and butterfat yield of dairy cows. Kungl. Lantb. Tidskrift 79 no 6 $\frac{1}{2}$:1-127.

- Lerner, J. Michael, Asmundson, V. S., and Cruden, Dorothy M. 1947. The improvement of New Hampshire fryers. *Poultry Science*. 26:515-524.
- Lush, Jay L. 1933. The bull index problem in the light of modern genetics. *Jour. Dairy Sci.* 16:501-522.
- _____ 1944. The optimum emphasis on dam's record when proving dairy sires. *Jour. Dairy Sci.* 27:937-951
- _____ 1945. Animal breeding plans, 3rd ed. Ames, Iowa, The Collegiate Press, Inc.
- _____ 1946. Chance as a cause of changes in gene frequency within pure breeds of livestock. *Amer. Nat.* 80:318-342.
- _____ 1947. Family merit and individual merit as basis for selection. *Amer. Nat.* 81:241-261 and 362-379.
- _____ 1948. The genetics of populations. Ames, Iowa, mimeographed.
- _____ 1949. Heritability of quantitative characters in farm animals. Eighth Int. Congress Genetics Proc. (*Hereditas*. Suppl. Vol.) 356-375.
- _____, Norton, H. W. III, and Arnold, Floyd. 1941. Effects which selection of dams may have on sire indexes. *Jour. Dairy Sci.* 24:695-721.
- _____ and Shultz, Earl N. 1936. Heritability of butterfat percentage and butterfat production in data with which sires have been proved in Iowa. (Abstract) *Jour. Dairy Sci.* 19:429-430.
- _____ and Straus, F. S. 1942. The heritability of butterfat production in dairy cattle. *Jour. Dairy Sci.* 25:975-982.
- Panase, V. G. 1946. An application of the discriminant function for selection in poultry. *Jour. Genetics* 47:242-248.

- Plum, M. 1934. Production in a large Jersey herd as affected by sires, dams, and yearly variation. Amer. Soc. Anim. Prod. Proc. 1933. 53-57.
- _____ 1935a. The relative importance of heredity and environment in determining the butterfat production of cows in Iowa Cow Testing Associations. Unpublished Ph.D. thesis, Ames, Iowa, Iowa State College Library.
- _____ 1935b. Causes of differences in butterfat production of cows in Iowa Cow Testing Associations. Jour. Dairy Sci. 18:811-825.
- Satterthwaite, F. E. 1946. An approximate distribution of estimates of variance components. Biometrics Bulletin 2:110-114.
- Seath, D. M. 1940. The intensity and kind of selection actually practiced in dairy herds. Jour. Dairy Sci. 23:931-951.
- _____ and Lush, Jay L. 1940. "Nicking" in dairy cattle. Jour. Dairy Sci. 23:103-113.
- Smith, H. Fairfield. 1936. A discriminant function for plant selection. Ann. Eugenics 7:240-258.
- Snedecor, G. W. 1946. Statistical methods. 4th ed. Ames, Iowa. The Collegiate Press, Inc.
- Tyler, W. J. and Hyatt, George Jr. 1947. The heritability of milk and butterfat production and percentage of butterfat in Ayrshire cattle. (Abstract) Jour. Anim. Sci. 6:479-480.
- Verna, Jean H. 1945. Degree of repeatability of production in the Iowa State College herd with respect to quantity of milk, quantity of fat and fat percent. Unpublished M. S. thesis. Ames, Iowa, Iowa State College Library.
- Winsor, C. P. and Clark, C. L. 1940. A statistical study of variation in the catch of Plankton nets. Jour. Marine Res. 3:1-34.

Wright, Sewall. 1921. Systems of mating. Genetics
6:111-178.

_____ 1934. The method of path coefficients. Ann.
of Math. Stat. 5:161-215.

_____ 1935. The analysis of variance and the
correlation between relatives with respect to
deviations from an optimum. Jour. Genetics.
30:243-256.

VIII. ACKNOWLEDGMENTS

A formal statement cannot express the depth of the author's indebtedness to Dr. Jay L. Lush for his suggestions and friendly counsel both during the course of this study and in the preparation of the manuscript. Dr. Lush suggested the topic and proposed solutions to many of the perplexing problems which arose. Sincere thanks also are expressed to Dr. L. N. Hazel for his valuable and willing assistance with the analysis and interpretation of the data.

The American Jersey Cattle Club kindly furnished the data and supported the statistical analysis. A special note of appreciation goes to Mr. John P. Beardsley, Director of Research, for his whole-hearted cooperation in making the original data available in such convenient form.

The author also wishes to express his appreciation for the opportunity to pursue graduate study under the direction of both Dr. Jay L. Lush and Dr. C. Y. Cannon. This association has been most stimulating and pleasurable.