

NOTE TO USERS

This reproduction is the best copy available.

UMI[®]

GENETIC VARIATION AND COVARIATION IN
TYPE AND BUTTERFAT PRODUCTION
AMONG JERSEY COWS

by

Walter Robert Harvey

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

Major Subjects: Animal Breeding
Dairy Husbandry

Approved:

Signature was redacted for privacy.

In charge of Major Work

Signature was redacted for privacy.

Head of Major Departments

Signature was redacted for privacy.

Dean of Graduate College

Iowa State College

1949

UMI Number: DP13270

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 DP13270

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

TABLE OF CONTENTS

	<u>Page</u>
I. INTRODUCTION.....	1
II. REVIEW OF LITERATURE.....	4
A. Heritability of Milk and Fat Production.....	4
B. Heritability of Type Ratings.....	9
C. Phenotypic Correlations Between Official Type Ratings and Fat Production.....	11
III. SOURCE AND ADJUSTMENT OF DATA.....	14
IV. THE INVESTIGATION.....	17
A. Problems of Estimation.....	17
1. Estimating heritability.....	19
2. Estimating genetic and environ- mental correlations.....	23
3. Eliminating the effects of years within herds.....	32
B. Daughter-Dam Analysis.....	38
1. Heritability of type.....	45
2. Heritability of production.....	52
3. Genetic correlation.....	54
4. Phenotypic correlations.....	55
C. Paternal Sister Analysis.....	57
D. Maternal Sister Analysis.....	68
E. Full Sister Analysis.....	75
F. Selection for Type and Production.....	83
V. DISCUSSION.....	92
VI. SUMMARY.....	99
VII. LITERATURE CITED.....	103
VIII. ACKNOWLEDGMENTS.....	106
IX. APPENDICES.....	107
Appendix A: Summary of the Number of Cows with Type and Production Records by Herds and States.....	108

	<u>Page</u>
Appendix B: An Algebraic Comparison of the Number of Sire Components in the Various Sums of Squares and Mean Squares with Proportionate and Disproportionate Frequencies.....	116

I. INTRODUCTION

Breeders of livestock are often confronted with two or more characters which appear to be positively correlated phenotypically. Human nature is such that the many times when the characters failed to go together are often forgotten and the few times when they do go together are remembered; consequently, many believe that those phenotypic correlations which do exist are much closer than have been found when thoroughly investigated. Only recently have the causes of some of the more important correlations between characters been studied. Although Galton introduced the correlation coefficient in 1888 it was not until Sewall Wright presented his method of path coefficients in 1921 that the underlying causes could be clearly defined and thus their effects more easily determined.

There are two different kinds of causes for a phenotypic correlation between two characters in the same animal, (1) one or more genes affect both characters alike, or (2) one or more of the variations in the environment which happen to that individual affect both characters in the same direction. In a breeding program aimed at improving two such characters it has been shown by Hazel (1943) that the amount that each of these causes contributes to the observed phenotypic correlation must be known if selection is to achieve maximum gain in both characters.

Contrary to the belief of many livestock authorities and breeders, a phenotypic correlation, regardless of its magnitude, does not of necessity indicate a common genetic cause. Therefore it follows that selection for one of two characters,

positively correlated in the individual, need not bring about an increase in the other character in the next generation. In fact, the opposite result will occur if the genetic correlation is of negative sign while the phenotypic correlation is positive.

It is generally agreed that type and production in dairy cattle are positively correlated phenotypically, although the exact magnitude of this correlation is somewhat uncertain. Quite likely it may vary for different conditions, e.g., the time the type rating was made with respect to the production records used, the number of different inspectors involved, the population of cows under consideration, and many others.

Several investigations concerning the inheritance of fat production have been made in recent years. Generally, it has been found that from 15 to 30 per cent of the intra-herd differences in single lactation records of production are accounted for by heredity. This fraction is termed the heritable fraction of the variance or, more generally, "heritability".

On the other hand, little is known regarding the heritability of official type ratings or the genetic correlation between type and production. Since estimates of these parameters are required in order to construct a selection index that will maximize the amount of gain in both characters, the present study was designed primarily to obtain such estimates. The type of analysis required to obtain estimates of these two parameters automatically gives estimates of several other important parameters, e.g., heritability of butterfat production, phenotypic correlation between type and production, environmental correlation

between type and production, the herd component for type and production, and several phenotypic correlations between relatives.

II. REVIEW OF LITERATURE

A. Heritability of Milk or Fat Production

Milk and butterfat production are of such major importance to the dairy industry that many investigators have studied their inheritance. Only those studies which are similar to the present study in scope of data and methods of analysis are reviewed here.

In an extensive investigation with Holstein-Friesian Advanced Registry records from 74 different herds Gowen (summarized in his book "Milk Secretion", 1924) used biometric methods to measure the resemblance between relatives for milk production. The correlations obtained between dam and daughter, full sisters, paternal half sisters, and maternal half sisters are given below.

	r	<u>No. in Study</u>
Dam and Daughter	$.497 \pm .021$	611
Full Sisters	$.548 \pm .027$	302
Paternal Half Sisters	$.362 \pm .015$	1400
Maternal Half Sisters	$.381 \pm .033$	498

Since these are the gross correlations obtained with no herd differences removed, they can not be compared directly with similar correlations obtained on an intra-herd basis. Gowen shows in his Chapter 19. that the differences between herds were large. In a later study with Jersey Register of Merit records Gowen (1934) estimated that from one-half to three-fourths of the variation in milk yield was due to heredity, but here also he did not discount the differences between herds. Inasmuch as some of the differences between herds are environmental and contribute to the gross correlations between relatives, Gowen's estimates are too high. If such herd-to-herd differences make up about

20 per cent of the total variance, as has been estimated by Plum (1935a), then Gowen's estimates of the heritable fraction of the variance are somewhere near .4 too large.

In a study with 2,394 daughter-dam pairs used in proving sires in Iowa Cow Testing Associations, Plum (1935a) reports an intra-herd correlation of .18 between the butterfat production of daughter and dam. In a more extensive study with 683 daughter-dam pairs scattered over 81 C.T.A. herds and reported in the same paper, Plum gives a corresponding correlation of .06. In the latter case only the first record available on production was used; whereas the former estimate was based on the highest record. Of the 683 daughter-dam pairs 246 had records starting not more than 3 months apart. The intra-herd correlation in this select sample was .10 between the first available records of daughter and dam and .27 between their nearly contemporary records.

Using 5,860 records made by 2,316 purebred and grade Guernsey, Holstein, and Jersey cows and distributed over an eleven-year period, Plum (1935b) found that the variance within year and herd was 10 per cent less than the variance within herd. The effect of general differences between years in the whole material was found to account for 2.8 per cent of the total variance. Using the first record available on 1,359 of these cows which were by 608 different sires, he estimated that the sire accounted for 19.6 per cent of the intra-herd variance. This would be very near the correlation between paternal half sisters in his data unless the number of daughters per sire was highly

variable. Plum points out that such an estimate is inflated since daughters of the same sire are often contemporaries.

In an earlier study with considerably less data on one Jersey herd with 183 cows, but with records distributed over a 17-year period, Plum (1934) found that the sire accounted for 22.6 per cent of the intra-herd variance. Since the daughters of a sire are more contemporary than all cows the effect of the sire also contained much of the time changes in the general herd environment. Register of Merit fat production records were used exclusively in this study. An intra-sire correlation of .20 between daughter and dam was reported with 158 daughter-dam pairs available.

Lush, Norton, and Arnold (1941) estimated that from 28 to 30 per cent of the observed differences in fat production within herd were transmissible from dam to daughter. The study was made on 676 daughter-dam pairs used to prove 103 sires in Iowa Dairy Herd Improvement Associations plus 3,010 daughter-dam pairs obtained from the first 8 volumes of the Holstein-Friesian Herd Improvement Registry yearbook. In the larger study there were 209 sires, with none having less than 6 daughter-dam pairs. The estimates of heritability were obtained by dividing the dams into two equal groups, a high half and a low half, two different times. The first division was made on the basis of their first record, and the second on the basis of their second record. Twice the ratio of the difference between the daughters of the dams in the high half and the daughters of the dams in the low half, to the difference between the average production of the high dams and the low dams was used as the estimate of heritability.

In an extensive investigation of the fat production records of the Swedish Red and White Cattle from 12 herds that were all well managed, Johansson and Hansson (1940) studied the resemblance between relatives using the first lactation of 300 days adjusted for age and current calving interval. The intra-herd correlations and the number on which each is based are given below.

	<u>r</u>	<u>No.</u>
Daughter-dam	.186	1,489 Pairs
Full Sisters	.145	235 Pairs
Maternal Sisters	.062	438 Pairs
Paternal Sisters	.120	1,557 Cows by 73 Sires

When only contemporary records of 621 daughters and dams were used with no adjustment for age or calving interval the intra-herd correlation increased to .211. These authors state that genetic differences in milk or butterfat yield are not manifested in an unfavorable environment, and thus the resemblances among relatives would increase as the environment improved. This supposition has not as yet been supported by evidence, although the problem needs much further investigation.

Lush and Straus (1942) studied the 305-day yields of fat in 2,154 daughter-dam comparisons used in proving 283 sires in Iowa D.H.I.A. during the period January 1, 1936 to December 31, 1939. They obtained an intra-sire regression of daughter on dam of .087 when converted to a single-record basis. When this regression was calculated intra-sire and intra-herd it dropped to .070, since some sires had daughters in more than one herd. Within sire groups the daughters were slightly less variable than their dams (a mean square of 3,838 as compared to 3,868). The dams were repeated if they had more than one daughter and no

adjustment was made for number of records. The dams had an average of 3.15 records each and the daughters averaged 1.68 records each, which should make the dams' records about .78 as variable as the daughters. Repeating the dams when they have more than one daughter also will bias the estimate of the dams' variance downward. On the other hand, the daughters within sire groups were more closely related than the dams. This should decrease the variance among the daughters.

Berry (1945) studied a selected group of 454 Holstein-Friesian Advanced Registry cows with at least 6 records each, which had 954 dams or daughters with records among which 661 dams or daughters had at least two records. All possible correlations between the records of the dam and daughter were calculated on both a total and an intra-herd basis. The intra-herd correlations varied from $-.004$ to $.128$, with an average of $.07$. This gives an estimate of heritability of 14 per cent for the heritability of single records of fat production from these data. Berry suggests that this estimate is about 6 per cent too low for use in a random bred population, since the cows studied were a highly selected group.

Recently Tyler and Hyatt (1947) studied 6,888 daughter-dam pairs in the Ayrshire breed. They estimated that differences in single records of milk and fat production were 31 and 28 per cent heritable respectively. These estimates were double the intra-sire regression of daughter on dam. The 6,888 daughters were by only 374 sires, indicating that the daughter dam pairs studied may have been those used to prove these sires even though this is

not stated in the brief report. If some of these sires had daughters in more than one herd, as would be expected, the heritability estimates which they give would be too large.

B. Heritability of Type Ratings

In a study of yearly unofficial classification ratings given by highly competent judges to cows and heifers in the Iowa State College Holstein herd, Johnson and Lush (1942) report a repeatability of .34 when ratings made at 10 months or less were omitted. This means that 66 per cent of the differences in yearly type ratings of the same cow were attributed to temporary environmental effects or to differences in ideals of the judges doing the classifying.

Hyatt and Tyler (1948) in a study of type ratings made by official inspectors at intervals of about 4 months on a herd of Ayrshire cows, estimated that repeatability of ratings made after freshening on the same cow was .55. In a later study of the same herd, Hyatt, Tyler, and Conklin (1949) report a correlation of .30 between any classification rating of a heifer before calving and a rating made after calving.

Tyler and Hyatt (1948) studied the inheritance of official type classification ratings using 3,738 Ayrshire cows sired by 368 bulls and 1,601 daughter-dam pairs. The intra-sire regression of daughter on dam, when both were classified by the same judge on the same day, was $.14 \pm .03$. In the paternal sister study, only those sires which had 6 or more daughters classified on the same day were included. The correlation between paternal

half sisters was found to be $.12 \pm .04$. They point out that at least some of this correlation would be expected to be caused by common environment of the half sisters. Then, assuming that daughters of the same sire were 30 per cent related instead of 25 per cent, since mates of a sire are often related, they state (P.66) that "heritability of type ratings could be estimated to be somewhat less than $0.\overset{1}{30} \times 0.12$ or 0.40 from these data". Apparently they assume this to be an unbiased estimate of heritability, regardless of the environmental contribution they mentioned since they later combine the .40 with the .28 (as estimated by doubling the intra-sire regression of daughter on dam) to obtain an estimate of .30 on which they give fiducial limits. Until the environmental contribution to a paternal half sib correlation can at least be approximated, and discounted, it seems certain that estimates of heritability from such correlations will almost always be too high.

In a study of unofficial type ratings on Holsteins made by members of the Animal Breeding Subsection of Iowa State College, Touchberry (1948) found no correlation between the type ratings of daughter and dam. However he had only 187 daughter-dam pairs for this study. He also estimated that the genetic correlation between type and production was zero. Significant cross correlations between one character in one relative and the other character in the other relative were found (.158 and -.186), but no explanation was given biologically for one being negative and the other positive. One would expect these two values to be simply two independent estimates of the same correlation.

C. Phenotypic Correlations Between Official Type Ratings and Fat Production

In an extensive investigation with 1,674 Register of Merit Jersey cows and scores on the same cows by 140 different judges, Gowen (1920) obtained a correlation of .19 between overall score and milk production.

In a later study with the same kind of data Gowen (1921) computed correlations between the types of cows, as scored by 19 different judges, and their 365-day milk production. The coefficients varied from -.10 to .64 for the different judges, with an average of .25. The data analyzed by Gowen were from many herds over many states. Thus, herd differences are present in the data and tend to make the correlations higher than they would be on an intra-herd basis if the herd averages in type and production were more closely related than the phenotypes of the same individual as reported by Tyler and Hyatt (1948).

In a report by Copeland (1938) on an investigation concerned with conformation, from the standpoint of the herd classification system and Jersey Register of Merit records, classification alone was not of much value for estimating producing ability. Specific information on the reliability of such an estimate is lacking since no statistical analyses were conducted.

In a more recent study on Jerseys Copeland (1941) noted considerable bias due to culling prior to classification and to more extensive testing of cows with good conformation. Another source of bias in his estimates of the phenotypic correlations

of type and production would be herd-to-herd differences, mostly environmental or caused by differences between the judges which tend to make the correlations higher than if the data were analyzed on an intra-herd basis. By using all cows (139) classified in the "fair" class and having records and an equal number in all higher classes being drawn at random, a correlation of .25 was obtained. When the data from only 68 herds, where Herd Improvement Registry testing had been practiced for at least one year prior to classification, were analyzed in the same manner with 76 in each class, the correlation was .31. From this information he concluded (1) that a high conformation score is not a reliable guarantee of high producing ability and (2) that the breeder must therefore continue to select for both type and production.

The fat production records of cows classified under the herd classification program of the American Jersey Cattle Club have been averaged separately (table 1) for each grade. These

Table 1. Average Production of Jersey Cows Classified for Type to April 1, 1946

Classification group	Number : records	Average Pounds : fat
Excellent	801	483
Very Good	4,213	460
Good Plus	6,060	448
Good	2,700	434
Fair	369	420

records were converted to the two-time, 305-day mature equivalent basis. The averages of the different classes indicate that a positive correlation does exist. The regression of production

on type rating is about 15.2 pounds of butterfat for each increase in grade.

Tyler and Hyatt (1948) calculated the phenotypic correlations between official type ratings and fat production records among 5,177 Ayrshire cows from 304 different herds. They found no significant difference between intra-herd correlation coefficients computed between (1) first record and type (.16), (2) nearest record to classification and type (.16), or (3) average of all records and type (.19). The latter correlation coefficient becomes .14 when reduced to a single-record basis, since the cows studied had an average of 3.5 records each. The correlation between the average type and average production for a given herd was .50, pointing out that type and production are more closely related as between herd averages than as concerns individual cows. The intra-herd regression of butterfat on type rating was 12.9 pounds per type interval when the average of all the records a cow had were considered.

III. SOURCE AND ADJUSTMENT OF DATA

Butterfat production records and type classification scores on 8,464 cows from 245 herds were obtained from the American Jersey Cattle Club. These included all cows with an official type rating from herds that were on Herd Improvement Registry test for at least four of the five years 1943 to 1947, inclusive. All classified cows with one or more records of production in that period were included. In Appendix A the herds are listed by states and for each herd is stated the number of cows on which the required data were available. Although the bulk of the herds are located in the Northeastern and Midwestern states, 39 different states are represented.

Each of the 245 herds had been classified at least once. The date of classification or the name of the inspector doing the classifying were not noted. Undoubtedly many of the cows were classified for type before 1943. The rules of the classification plan provide that if a cow is classified more than once only the highest classification she has received is used. Which of the cows in these data had had their type rating raised on reclassification was not noted. It is believed that this would be a rather small percentage of the total. From the very beginning of the Herd Classification program the American Jersey Cattle Club has required that all eligible animals in the herd be classified, rather than a select few, whenever a breeder enrolls in the program.

Under the rules of the classification program the cow is classified into one of six grades; "Excellent", "Very Good",

"Good Plus", "Good", "Fair", or "Poor". No "Poor" cows were found in the data studied here. For the analysis the grades were given consecutive numerical scores beginning with 2 for a "Fair" cow through 6 for an "Excellent" cow.

Since 1943 the American Jersey Cattle Club has converted all official records (either in Register of Merit or in Herd Improvement Registry) to the twice-a-day milking, mature equivalent basis. If the actual record was for more than 305 days, only the first 305 days are used by the club. If the record was for less than 305 days, but was recorded as a complete lactation for that cow, no adjustment for length was made. Incomplete records of less than 305 days are not recorded by the club, unless the breeder so desires. The conversion factors used to convert all records to the mature equivalent basis are shown in table 2 below. The factors which were used to convert 3X and 4X records to a 2X milking frequency are shown below table 2.

Table 2. Conversion Factors Used by A.J.C.C. to Convert Records to the 2X, Mature Equivalent Basis.

Age at Calving	Factor	Age at Calving	Factor
Under 2.0	1.36	5-	1.01
2.0 - 2.5	1.26	6-	1.00
2.6 - 2.11	1.19	7-	1.01
3.0 - 3.5	1.13	8-	1.04
3.6 - 3.11	1.08	9-	1.08
4.0 - 4.5	1.05	10 & up	1.12
4.6 - 4.11	1.03		

Corrections for Frequency of Milking

3X to 2X = .833

4X to 2X = .741

No record of less than 270 days duration was used in the present analysis. Since the increased variation caused by using the few records that were shorter than 305 days was expected to be less than one half of one per cent of the intra-herd variance, no adjustment for length of record was made. If a record of less than 305 days does include all of that lactation for a given cow it is doubtful that a correction for length is justified on genetic grounds. Such an adjustment assumes the cow actually had the genetic ability to produce for 305 days, but did not simply because of the environmental conditions. If such an assumption is not wholly true then some of the truly genetic variability would be removed by adjusting records to 305 days.

IV. THE INVESTIGATION

A. Problems of Estimation

Table 3 shows a sample of the original records as first received from the American Jersey Cattle Club on mimeographed data sheets. Altogether there were 23,330 records, 12,405 cows and 293 herds. Of these, 3,941 of the cows and 48 of the herds did not have type ratings, but were analyzed along with all those with type, in a separate study involving only production. After

Table 3. Sample of Data as Obtained from A.J.C.C.

Herd No.	Type Rating	Cow No.	Sire No.	Dam No.	Year Record Ended					M.E. Prod.
					'43	'44	'45	'46	'47	
140504	5	1428580	401061	1219229				*		447
140504	5	1428581	401061	1070231				*		411
140504	4	1428582	412335	1303072				*		446
140504	5	1315832	401061	1116514		*				403
140504	5	1315832	401061	1116514				*		401

receiving the data in this form they were all then transferred to International Business Machine cards. One card was made for each lactation of the cow.

In order to remove the effect of year within herd and to obtain estimates of the intra-year variances for each herd, each record of fat production was expressed as a plus or minus deviation from the average of that herd in that year. A product moment correlation was computed between the average M. E. production of the herd and the intra-year variance of that herd. This correlation was found to be $.46 \pm .05$. A scatter diagram of the means and variances is presented in figure 1. The diagram indicates that a slightly curvilinear relationship exists between the two variables.

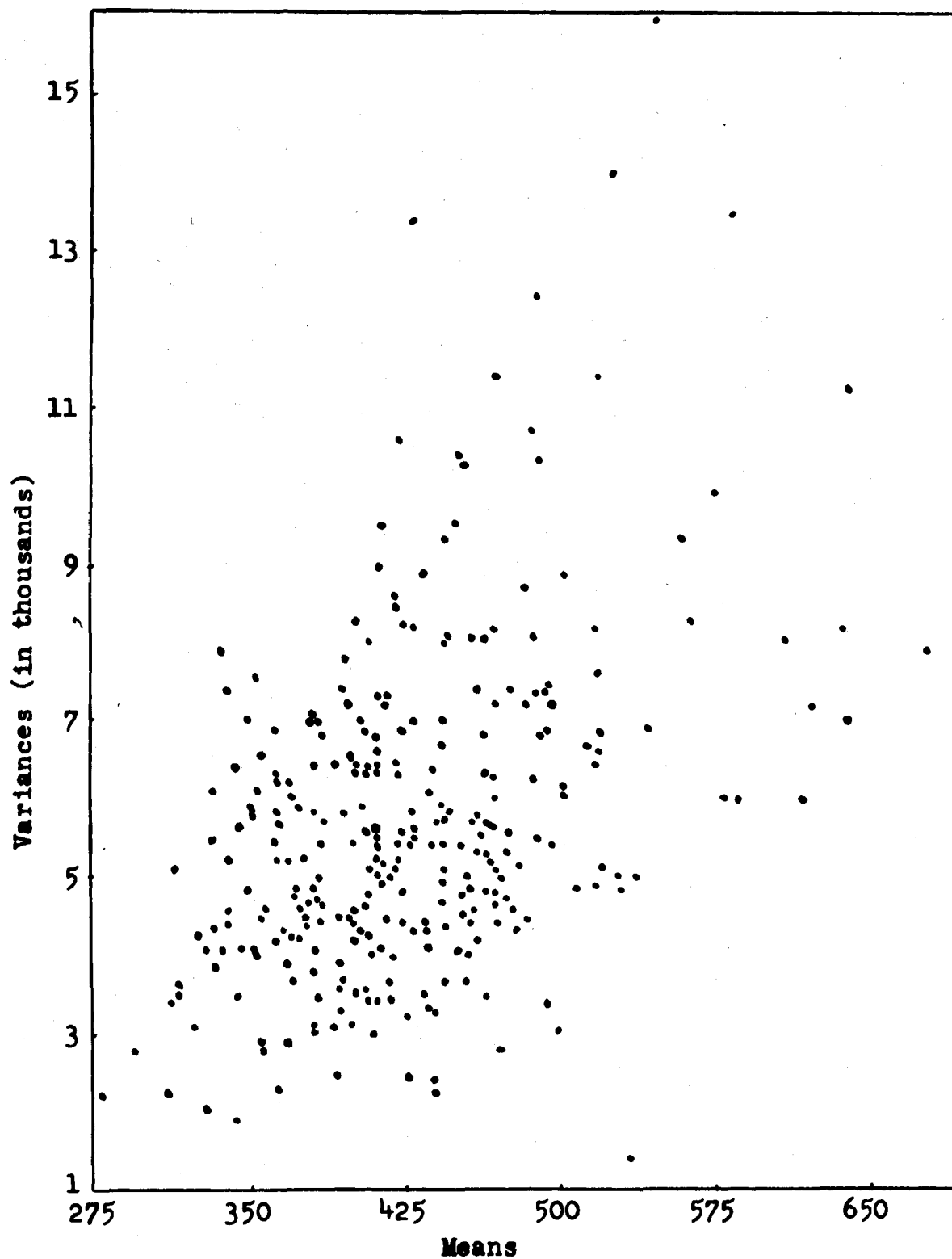


Figure 1. Scatter Diagram Showing the Relation Between the Herd Mean Fat Yield and the Intra-herd Intra-year Variance for 293 Herds.

The wide variation in the variances suggested that a test for heterogeneity of variance be made. Bartlett's test, as outlined by Snedecor (1946), yielded a highly significant χ^2 .

An effort was then made to divide the herds into 9 different groups on the basis of the herd average. The first group consisted of all herds with a mean of 330 pounds of fat or less and the last group consisted of herds with means of 571 pounds or more. The groups intervening were at 30 pound intervals. Bartlett's test still gave highly significant χ^2 's within all groups.

The effects of heterogeneity of variances between subgroups in a population on estimates of variance or of covariance components are not fully understood. Intuitively one would suppose such estimates would have greater sampling errors than those obtained from a population where the variance was homogeneous. This would not necessarily mean that a ratio of components, estimated from a population with heterogeneous variance, would be biased. Satterthwaite (1946) has pointed out that in many problems, which involve estimation of components, the assumptions regarding homogeneity of variance can be ignored.

1. Estimating heritability

Assuming that the genetic and environmental effects are linear and not correlated with each other, the total phenotypic or observed variation has been written

$$\sigma_X^2 = \sigma_G^2 + \sigma_I^2 + \sigma_D^2 + \sigma_E^2$$

Where: σ_x^2 = total variance
 σ_G^2 = additively genetic variance
 σ_I^2 = variance due to epistatic deviations
 σ_D^2 = variance due to dominance deviations
 σ_E^2 = environmental variance

The last term in this equation (σ_E^2) consist of all environmental effects, including both permanent (σ_{PE}^2) and temporary (σ_{TE}^2) environmental effects. Although σ_D^2 and σ_I^2 are genetic in the sense that they are variations caused by combinations of genes, such variations are not transmissible from one generation to the next---except for a small amount of σ_I^2 . The combinations segregate and such effects as were peculiar to the combination rather than to the constituent individual genes, may not reappear in the offspring.

Dominance, epistasis, and genic effects are so defined that no correlations exist between them. A correlation between the sum of these three effects (H) and the environmental effects (E) will occur in the total population if breeders who have the better cows also give them better care and management than the average. By removing the part of σ_E^2 that is due to differences between herds one automatically removes σ_{HE}^2 caused by one breeder giving his cows better management than another breeder does. If a breeder gives the cows that have better than average heredity better care and management than the other cows in his herd, H and E will be correlated even though herd differences are removed. Since the heredity of a cow must be estimated from her phenotype or the phenotypes of relatives, enough mistakes will

be made so that r_{HE} within herd is likely to be small, even though the breeder might try intentionally to give the best cows the better environment and vice versa.

Lush (1940) defines heritability in the "narrow" sense as that fraction of the total phenotypic variance that is additively genetic, i.e.,

$$\text{Heritability} = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_I^2 + \sigma_D^2 + \sigma_E^2 + \sigma_{EH}^2}$$

Twice the intra-sire regression of offspring on dam has been found to be very useful in estimating this fraction in outbred populations, since both numerator and denominator come closer to containing the variances as indicated in the equation than do estimates from other sources. However, in dairy data where a sire has daughters in more than one herd, the heritability estimate thus obtained will be too high to the extent that some of the σ_E^2 in the herd component will remain in both numerator and denominator. On the other hand, the intra-sire and intra-herd regression of offspring on dam does remove most of the differences in the environment that are caused by time trends and may be similar for daughter and dam but different from one daughter-dam pair to another. This has some advantage over the intra-herd regression of offspring on dam when the data used extend over several years and a time trend in average production is present.

In dairy data the regression of offspring on dam is generally preferred to the correlation since the dams are likely to

be less variable than the daughters because of selection. The parameter to be estimated is the fraction of the total or phenotypic variance in the dams that is transmissible to the progeny¹; consequently, the correlation will give an underestimate if the dams are a selected group.

Estimates of heritability made from paternal or maternal half sister correlations have greater sampling errors than those made from daughter-dam comparisons or full sister correlations, since the former are multiplied by four instead of two. Estimates of heritability from half sister correlations are scarce in data concerning type or production of dairy cattle. Another disadvantage of using the phenotypic resemblance between half sisters is the difficulty of discounting the environmental correlations. This is particularly important with paternal half sisters since most of the daughters of a given sire are born within the short space of a year or two.

When heritability is estimated from an average of n records for each cow, the estimate should be converted to the basis of a single record, so that it could be compared simply with results from other studies. Lush and Straus (1942) have pointed out that a correlation between X and Y , where both X and Y are averages of n records, is $\frac{n}{1 + (n-1)r}$ times as large as one based on single records. Here r is the correlation between the records of the

¹Actually the regression of the transmissible part of the genotype of the dam on the dam's phenotype is the parameter desired. Since the transmissible part of the dam's genotype cannot be observed directly, the next best thing available, the daughters of the dams, must be used in its place.

same cow, i.e., "repeatability". To adjust a regression of offspring on dam to a single-record basis, where the number of records per dam is variable, Lush and Straus (1942) give the following formula, obtained from Cochran.

$$b = b' \left[\frac{1 + (\bar{m}-1)r_{dd}}{\bar{m}} + \frac{\sigma_m^2(1-r_{dd})}{\bar{m}^3} \right]$$

Where: b = regression of daughter on dam with single records.

b' = regression of daughter on dam with averages.

\bar{m} = average number of records per dam.

σ_m^2 = variance of m .

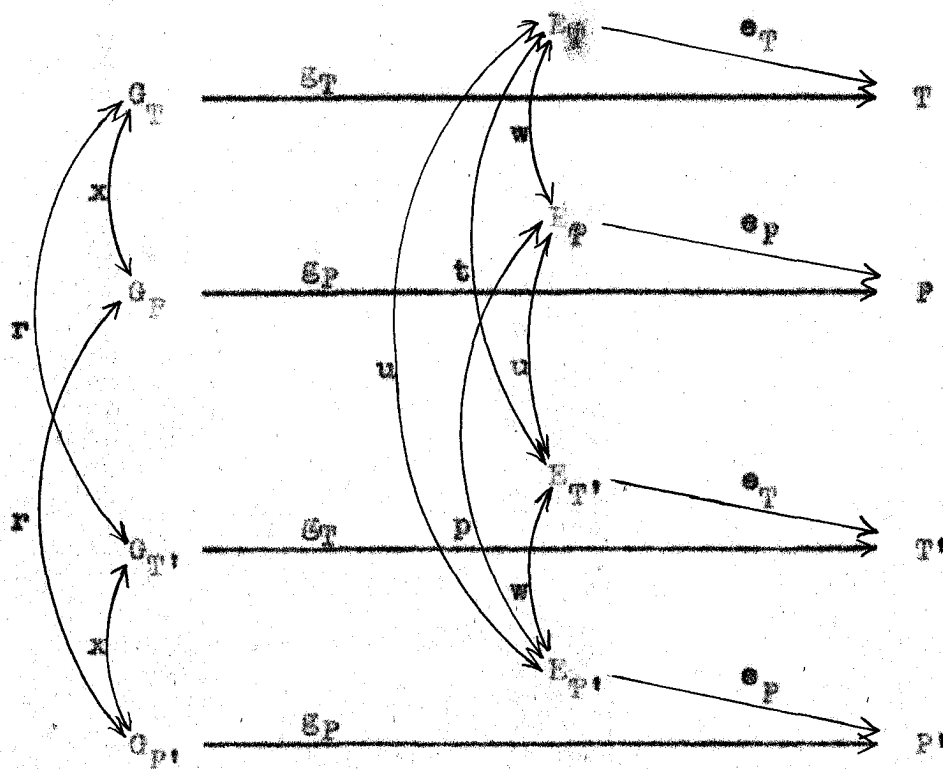
r_{dd} = repeatability of production records.

The last term in this formula can usually be ignored for practical purposes since little error will result unless σ_m^2 is quite large.

2. Estimating genetic and environmental correlations

Using the method of path coefficients as introduced by Wright (1921), Hazel (1941) has derived the formula for obtaining estimates of genetic correlations. The genetic and environmental components of the biometric relations between relatives for two characters, such as type and production, are diagrammed in figure 2. The derivation of the formula for estimating \underline{x} is shown below the figure. An estimate of \underline{x} is unbiased only if \underline{t} , \underline{p} , and \underline{u} each equal zero or if their effects are known and can be discounted. The bias caused by these environmental correlations not being zero will be discussed later.

The phenotypic correlations r_{tp} and r_{pp} each consist of



x = genetic correlation between type and production.

r = genetic relationship of the relatives concerned.

$h_T = \frac{\sigma_{G_T}}{\sigma_T}$ = square root of heritability of type.

$h_P = \frac{\sigma_{G_P}}{\sigma_P}$ = square root of heritability of production.

When $t = p = u = 0$

$$r_{TT'} = h_T^2 \quad r_{PP'} = h_P^2$$

$$r_{TP'} = r_{PT'} = h_T r x h_P$$

and x can be estimated as $\frac{r_{TP'}}{h_T h_P r}$

$$\text{or } x^2 = \frac{r_{TP'} r_{PT'}}{r_{TT'} r_{PP'}} = \frac{b_{PP'} b_{PT'}}{b_{TT'} b_{PP'}}$$

Figure 2. Biometric Relations Between the Genotypes (G_T and G_P), Environments (E_T and E_P), and Phenotypes (T and P) of two Relatives When the Characters, Type and Production, are Involved.

an environmental and a genetic component, i.e.,

$$r_{TT'} = g_T^2 r + e_T^2 t$$

and

$$r_{PP'} = g_P^2 r + e_P^2 p$$

Assuming that g_T^2 and g_P^2 as obtained from the intra-herd regression of daughter on dam describe accurately the heritability of type and production, one can obtain estimates of t and p in other relatives without difficulty. The accuracy of such estimates, in addition to sampling errors, depends on t and p actually having been zero between daughter and dam or their effects having been eliminated and the accuracy of the assumption that heritability is the same in both populations.

The situation becomes a little more complex when estimates of w and u are desired. The formulas for estimating these two environmental correlations can be derived from the formulas for the phenotypic correlations, r_{TP} , and $r_{TP'}$, as shown below.

$$r_{TP} = r_{T'P'} = g_T x g_P + e_T w e_P$$

hence

$$w = \frac{r_{TP} - g_T x g_P}{e_T e_P}$$

and

$$r_{TP'} = r_{PT'} = g_T r x g_P + e_T u e_P$$

hence

$$u = \frac{r_{TP'} - g_T r x g_P}{e_T e_P}$$

To estimate the environmental correlations, w and u , requires that one have estimates of (1) heritability of type (g_T^2), (2) heritability of production (g_P^2), and (3) the genetic correlation between type and production (x) from some other source.

Hazel (1941) has pointed out the desirability of using regressions of offspring on dam to estimate genetic correlations as contrasted to correlations between daughter and dam or to sire and dam components from half sib analyses. The primary advantages of such estimates are summarized below.

- (1) Only one Mendelian segregation separates daughter and dam, i.e., the relationship is as close as it can possibly be between any relatives in an outbred population, except for identical twins. For example, with the same number of degrees of freedom the heritability estimates from comparisons of daughter and dam have sampling errors half as large as those of estimates obtained from relatives separated by two Mendelian segregations.
- (2) In dairy data, where herd differences have been removed, the environmental correlation between daughter and dam can usually be assumed to be zero, unless the data extend over a long period of time and no attempt is made to remove differences between years.
- (3) The regression eliminates the necessity for the dams to be unselected and hence gives an unbiased estimate.

Hazel, Baker, and Reinmiller (1943) outlined the method for obtaining estimates of genetic correlations from resemblances between half sibs. They indicate that the covariance between sires (or between dams--if maternal half sisters are being studied) in an outbred population, contains a component of covariance

which, if all differences between sires are assumed to be entirely genic, is 1/4 of the additively genetic covariance for the two characters being studied. This can be demonstrated in the present study with type and production by first assuming an appropriate mathematical model for the two characters, and then determining the expected values of the necessary uncorrected cross-products. The validity of the proof depends of course on the mathematical model corresponding accurately to the actual facts.

The model that appears to describe most accurately the situation for type in the analysis of paternal sisters¹ is

$$T_{ijk} = \mu_T + h_{T_i} + \frac{g_{ST_{ij}}}{2} + \epsilon_{T_{ijk}}$$

where:

T_{ijk} = type rating of the k^{th} cow by the j^{th} sire in the i^{th} herd.

μ_T = general mean for type.

h_{T_i} = effect of the i^{th} herd on type.

$\frac{g_{ST_{ij}}}{2}$ = effect of the genotype of the j^{th} sire in the i^{th} herd on type.

$\epsilon_{T_{ijk}} = c_{T_{ijk}} + e_{T_{ijk}}$

and

$c_{T_{ijk}}$ = effect of the dam's genotype, the joint effects of the sire's and dam's genotypes, (all of the dominance and most of the epistatic effects), and permanent environmental effects on the type rating.

¹The models would be the same for the maternal sister analysis, except that $\frac{g_{DT_{ij}}}{2}$ and $\frac{g_{DP_{ij}}}{2}$ would be used in place of $\frac{g_{ST_{ij}}}{2}$ and

and $\frac{g_{SP_{ij}}}{2}$, respectively.

e_{Tijk} = random error--assumed to be normally distributed with variance σ_e^2 and a mean of zero. It includes all the temporary environment effects or judge differences that would make a cow's type rating different from one time to another.

The model for the average fat production of a given cow is

$$P_{ijk} = \mu_P + h_{P1} + \frac{g_{SP1j}}{2} + \epsilon_{Pijk}$$

where:

P_{ijk} = average production of n records of the k^{th} cow by the j^{th} sire in the i^{th} herd.

μ_P = general mean for production.

h_{P1} = effect of the i^{th} herd on production.

$\frac{g_{SP1j}}{2}$ = effect of the genotype of the j^{th} sire in the i^{th} herd on production.

$\epsilon_{Pijk} = c_{Pijk} + \frac{\sum [\text{error associated with each record}]}{\text{no. of records for that cow}}$

and

c_{Pijk} = effect of the dam's genotype, the joint effects of the sire's and dam's genotype, and permanent environmental effects on fat production.

h_{T1} and h_{P1} are the average effects of the i^{th} herd on type and production expressed as a deviation from zero. That is, if all herds had the same effect there would be no component of variance from this source.

$\frac{g_{ST1j}}{2}$ and $\frac{g_{SP1j}}{2}$ are the average "true" effects of the j^{th} sire in the i^{th} herd. In statistical terminology this means that the expected value of the effect of the j^{th} sire in the i^{th}

herd is the same for each of his k daughters. If the models had been set up so that the sires were classified only by sires rather than sires within herd, some of the herd differences from the crossproduct between sires would have come out in the crossproduct within sires, since some of the sires had daughters in more than one herd. In estimating components of variance and covariance to be applied to intra-herd conditions it is particularly important to eliminate the herd effects from the mean squares or covariances for between sires within herd and within sires.

In all studies where samples are studied and the results are generalized, it is necessary to assume that the data studied were representative of the more general population. Although this is generally recognized, its importance is often overlooked. Results should not be generalized if selection of the sample studied has occurred and the effects of such selection cannot be discounted. Since the herds used in the present study were required to have been on Herd Improvement Registry test for at least four out of five years and each of them had cows which had been officially classified for type, it is obvious that the results obtained may not be applicable to a population of herds where some or all of them do not meet these requirements.

With these two models at hand one can easily determine the expected or average values of the uncorrected crossproducts and the correction term, in terms of the components which they contain. From these the theoretical composition of the covariances can

be obtained. As seen from the two models given above, the data are classified according to herds and then by sires within herd; hence, the crossproducts required are (1) total, (2) between sires, and (3) between herds. From these three crossproducts and the correction term can be obtained the corrected crossproducts between sires within herds and within sires. The expected composition of each of the three uncorrected crossproducts and of the correction term are given below.

Total Crossproduct

$$E\left\{\sum_{ijk} T_{ijk} P_{ijk}\right\} = n... \left[\mu_T \mu_P + \text{Cov}_H(TP) + \frac{\text{Cov}(G_T G_P)}{4} + \text{Cov}_E(TP) \right]$$

Crossproduct Between Sires

$$E\left\{\sum_{ij} \frac{T_{ij.} P_{ij.}}{n_{ij.}}\right\} = n... \mu_T \mu_P + n... \text{Cov}_H(TP) + n... \frac{\text{Cov}(G_T G_P)}{4} + \sum_{ij} n_{ij} \text{Cov}_E(TP)$$

Crossproduct Between Herds

$$E\left\{\sum_i \frac{T_{i..} P_{i..}}{n_{i..}}\right\} = n... \mu_T \mu_P + n... \text{Cov}_H(TP) + \sum_i \frac{\sum_{ij}^2 n_{ij.}}{n_{i..}} \frac{\text{Cov}(G_T G_P)}{4} + \sum_i n_i \text{Cov}_E(TP)$$

Correction Term

$$E\left\{\frac{T... P...}{n...}\right\} = n... \mu_T \mu_P + \frac{\sum_i^2 n_{i..}}{n...} \text{Cov}_H(TP) + \frac{\sum_{ij}^2 n_{ij.}}{n...} \frac{\text{Cov}(G_T G_P)}{4} + \text{Cov}_E(TP)$$

A dot signifies summation over that subscript. $\text{Cov}_H(\text{TP})$ is the covariance between the average type and average production of each herd. $\text{Cov}_E(\text{TP})$ contains $3/4$ of the additively genetic covariance plus an entire component of environmental covariance between type and production in the same individual.

If the environmental effects for type in one paternal sister are correlated with the environmental effects for production in another sister, $\text{Cov}(G_T G_P)$ will be overestimated in this analysis. This is true because $\text{Cov}(G_T G_P)$ can only be estimated from the phenotypic covariance between type in one relative and production in another, i.e.,

$$\text{Cov}(\text{TP}') = \text{Cov}(\text{PT}') = \frac{\text{Cov}(G_T G_P)}{4} + \text{Cov}(E_T E_{P'})$$

Likewise the estimates of $\sigma_{G_T}^2$ and $\sigma_{G_P}^2$ from the paternal sister analysis will contain an environmental component if daughters of a sire are reared under similar environment which is different from that of daughters of other sires in the same herd, so that the genetic correlation,

$$\frac{.25 \text{Cov}(G_T G_P) + \text{Cov}(E_T E_{P'})}{\sqrt{[.25 \sigma_{G_T}^2 + \text{Cov}(E_T E_{T'})][.25 \sigma_{G_P}^2 + \text{Cov}(E_P E_{P'})]}}$$

is likely to be biased unless by chance the ratio of the environmental contribution to the numerator and denominator is the same as the actual genetic correlation. The bias may be in either direction depending on the size of \underline{t} , \underline{p} , and \underline{u} and the actual

magnitude of the genetic correlation between type and production.

3. Eliminating the effects of years within herds.

As indicated in table 3 on page 17, the data on production were distributed over a five year period. Since daughters and dams will have most of their records in different years, any effects of environmental differences between years within herd would bias the regression of daughter on dam downward if such effects are not eliminated. On the other hand, bulls tend to sire most of their daughters in one or two years only and consequently a larger portion of the yearly effects would be expected to appear in the sum of squares between sires than in the sum of squares within sires. This would have the effect of biasing the sire component and making heritability appear too large when it is estimated from paternal sister resemblance. From these considerations it seemed necessary that the effects of years within herd be removed in all analyses. The importance of doing this depends, of course, on whether the effects of year-to-year variations within herd really are large.

At first it seemed that this problem could be solved most easily by expressing each record simply as a deviation from the yearly herd average and subsequently working only with such deviations. However, since not all the cows had a record in every year, the differences between years which would thus be removed actually contain some cow differences and those differences which would remain within cow would still contain some cow differences which could not be separated easily.

Since the effects of years and cows are confounded over all the data; i.e. are non-orthogonal¹, no sum of squares could be obtained in which only the cow, year, or error component was present. Dr. J. L. Lush suggested the analysis given in table 4, which enables one to obtain four separate equations from which estimates of these three components can be made by simultaneous solution.

Table 4. Analysis of Variance Used to Obtain Estimates of the Error (E), Cow (C), and Year (Y) Components

Source of Variation	d.f.	Composition of Mean Square
Within Herd	$\sum_1 (n_i - 1)$	
Between Years	$\sum_1 (r_i - 1)$	$E + \bar{c}_2 C + \bar{y}_2 Y$
Within Year	$\sum_1 (n_i - r_i)$	$E + C$
.....
Within Herd	$\sum_1 (n_i - 1)$	
Between Cows	$\sum_1 (k_i - 1)$	$E + \bar{c}_1 C + \bar{y}_1 Y$
Within Cows	$\sum_1 (n_i - k_i)$	$E + Y$

Where:

n_i = number of records in the i th herd
 r_i = " " years " " " "
 k_i = " " cows " " " "

¹Yates (1933, P. 109) gives the following definition of orthogonality; "that property of the design which ensures that the different classes of effects shall be capable of direct and separate estimation without entanglement".

The assumptions underlying the analysis given in table 4 become more apparent by considering the linear mathematical model on which it is based.

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

where:

Y_{ijkl} = the i^{th} record of the k^{th} cow in the j^{th} year in the i^{th} herd.

μ = general mean associated with all records.

h_i = effect of the i^{th} herd.

y_{ij} = effect of the j^{th} year in the i^{th} herd.

c_{ijk} = effect of the k^{th} cow in the j^{th} year in the i^{th} herd.

e_{ijkl} = random error associated with each individual record, assumed to be normally distributed with mean zero and variance σ_e^2 .

Since records were classified by years according to the date the record ended, it was possible for a cow to have two records recorded in the same year. This changes the composition of the mean squares within year and within cow slightly from that shown in table 4. Instead of the coefficients for the cow and year components being one in these two mean squares they would be a little less than one. The number of cows with two records recorded in the same year was believed to be small enough (less than 500) in these data that no serious error would result from ignoring this. It does have the effect of making the estimate of E slightly too small and the estimates of C and Y correspondingly too large.

As seen from the model above, this analysis places both what might be called "general year effect" and the "herd-year interaction" into the year within herd (Y) component. Since the herds being studied were scattered over most of the United States, one would expect the "herd-year interaction" to make up most of Y. This would be particularly true if individual changes in herd management or varying climatic conditions from one year to another were the main causes of yearly differences within herd. Management would vary from herd to herd and climatic conditions would be expected to vary considerably from one section of the country to another and would largely cancel each other in the average for the whole country each year. Effect of changes in the general price levels, whereby increased production might be much more profitable in some years than in others, would not cancel each other so much and would largely remain in the general year effect.

The E component contains the random temporary environmental effects which make a cow's record higher one time than another but are not the same for other cows in that herd the same years. The fact that such a component exists prevents the records made by the same cow from being perfectly repeatable, even though such records are adjusted for the differences which can be attributed to years within herd as well as for age, frequency of milking, and length of lactation.

From the model given above can be calculated the number of times each component of variance is expected to be present in all

the pertinent uncorrected sums of squares and in the correction term. The coefficients \bar{c}_1 , \bar{c}_2 , \bar{y}_1 , and \bar{y}_2 in table 4 can then be obtained. The composition of the uncorrected sums of squares and of the correction term are presented in formulas in table 5.

Table 5. Theoretical Composition of Uncorrected Sums of Squares and Correction Term. All Lactations.^a

Source of Variation	H	Y	C	E
Total	n...	n...	n...	n...
Between Herds	n...	$\sum_i \frac{\sum_j n_{ij}^2}{n_{i..}}$	$\sum_i \frac{\sum_k n_{i.k}^2}{n_{i..}}$	Number of Herds
Between Years	$\sum_j \frac{\sum_i n_{ij}^2}{n_{.j.}}$	$\sum_j \frac{\sum_i n_{ij}^2}{n_{.j.}}$	Number of Years	Number of Years
H X Y Subclass	n...	n...	Number of Herd-Year Subclasses	Number of Herd-Year Subclasses
Between Cows	n...	Number of Cows	n...	Number of Cows
Correction Term	$\frac{\sum_i n_{i..}^2}{n...}$	$\frac{\sum_{ij} n_{ij}^2}{n...}$	$\frac{\sum_k n_{i.k}^2}{n...}$	1

^a μ^2 has been omitted purposely, since it is eliminated from all mean squares.

The analysis as outlined above was carried out by Legates (1949) on all data received from the American Jersey Cattle Club. The estimate of Y was not made separately for the portion of those cows which had type ratings, but the number of year components in all mean squares for production among these cows was

calculated. Since there appears to be no logical reason for the year affects to be different among those cows with type ratings and among those not classified for type, the estimate of Y obtained by Legates was used here in obtaining estimates of other components. For more details of the analysis by which Y was evaluated, the reader is referred to Legates.

B. Daughter-Dam Analysis

All daughters and dams used in this analysis had an official type rating and one or more fat production records. Also the records of both daughter and dam had to be made in the same herd. There were 2,044 dams with 2,786 daughters which met these requirements. They were distributed among 226 herds. The average number of records per dam was 2.46, while the daughters averaged 1.89 records each.

The percentages of these dams, daughters, and all cows in each type class and the average fat production of each such group are presented in table 6. That the dams average a bit

Table 6. Comparison of Type Ratings and Fat Production of Dams, Daughters, and All Cows.

Type Rating	Percentage			Average Production		
	Dams	Daughters	All Cows	Dams	Daughters	All Cows
Excellent (6)	4.6	2.3	3.0	453.5	503.8	464.1
Very Good (5)	37.6	31.9	33.4	451.8	459.6	446.3
Good Plus (4)	43.3	48.1	45.0	434.4	441.8	431.7
Good (3)	12.8	16.1	16.8	414.7	420.8	409.9
Fair (2)	1.8	1.7	1.9	401.0	404.3	397.4
Means	4.30	4.17	4.19	438.7	444.8	433.3

higher in type than all cows may indicate that breeders who enrolled in the herd classification program have generally been retaining longer those cows with slightly better than average type.

The average production of the dams was less than the average of their daughters but greater than the average of all cows

which had a type rating. This seems to indicate that these breeders were also selecting cows on the basis of their production records as well as their type ratings. However, it seems that they were giving about twice as much attention to type, since the average for the dams in type was about $1/7$ of a standard deviation above the average of all cows as compared to a difference of about $1/15$ of a standard deviation in production.

More of the dams than of all cows had an "Excellent" type rating, yet the average production of these dams was less than the corresponding average for all cows. The reverse situation is true for the dams with "Good Plus", "Good", or "Fair" ratings. Although these differences are not statistically significant they do seem to indicate that these breeders were retaining for breeding purposes those cows which had either a high type rating or a high production record, even when they were undesirable in the other trait.

Since dams and their daughters are likely not to have made their records in the same years, it is expected that the year effects which influence a dam's records will usually not be identical with those which influence her daughter's records. In these data the dams made more of their records in the first two or three years than in the last two years, while more of the daughters' records were made in the latter two or three years. This would be expected simply because many of the cows with records in the latter years were not yet old enough to have daughters with records in the data. It was desired, therefore, to eliminate

the effects of years in both the intra-herd variance of the dams and the covariance between daughter and dam in order to obtain an intra-herd regression of daughter on dam that would be free of yearly effects.

The mathematical model used to describe the average production of the dams for the variance analysis was,

$$\frac{D_{i.k}}{n_{i.k}} = \mu + h_i + \sum_j \frac{n_{ijk} y_{ij}}{n_{i.k}} + e_{ik}$$

Where:

$D_{i.k}$ = sum of all records (over all years) of the k^{th} dam of the i^{th} herd.

$n_{i.k}$ = number of records for the k^{th} dam of the i^{th} herd.

μ = general mean, characteristic of the population of dams.

h_i = effect of the i^{th} herd.

n_{ijk} = number of records for the k^{th} dam in the j^{th} year in the i^{th} herd. In most cases this would be 1 or 0.

y_{ij} = effect of the j^{th} year in the i^{th} herd.

$\sum_j \frac{n_{ijk} y_{ij}}{n_{i.k}}$ = yearly effect remaining in the average of all records of the k^{th} dam in the i^{th} herd.

$$e_{ik} = c_{ik} + \sum_j \frac{e_{ijk}}{n_{i.k}}$$

and

c_{ik} = effect of the k^{th} dam in the i^{th} herd.

e_{ijk} = random error associated with each record. Assumed $N(0, \sigma_e^2)$, i.e., normally distributed with mean zero and variance σ_e^2 .

The year effect remaining in the average of $n_{i.k}$ records will be $1/n_{i.k}$ of each of the year effects associated with the years within herd in which the k^{th} dam had her records. Therefore, in order to determine the number of year components in the uncorrected sum of squares it is necessary to know the distribution of all records with respect to years by dams and by herds. The theoretical composition of the uncorrected sums of squares in the total and between the herds and in the correction term is given in table 7. After one has calculated the expected number of the herd, year, and error components in the uncorrected sums of squares and in the correction term, the expected number of these components in the mean squares can easily be determined by the conventional method used to obtain the sums of squares and mean squares.

By using the average production of the dams the effect of the dam's genotype¹ is confounded with permanent and approximately $1/n_{i.k}$ of the temporary environmental effects and with whatever dominance and epistatic deviations are peculiar to that dam. These effects all come out together in the E component. The E component for averages is therefore different from the one with individual records (tables 4 and 5), since the latter contains only the temporary environmental effects, while the E component

¹As used here the "dam's genotype" includes only the additive effects plus a small amount of the epistatic deviations, i.e., those which are transmissible from parent to offspring. It may be thought of as the "cow effect" (c_{ik}) minus permanent environmental effects and minus the dominance and most of the epistatic effects on that cow.

in table 7 contains $\frac{1}{n_{1.k}}$ of that plus all the effect of real differences between cows in the same herd.

The covariance between the dam's genotype and the daughter's genotype is required to estimate heritability. Therefore, it is

Table 7. Theoretical Composition of Uncorrected Sums of Squares and Correction Term. Average Fat Production of Dams.

Source of Variation	H	Y	E
Total	$n_{..}$	$\sum_{ijk} \left(\frac{n_{ijk}}{n_{1.k}} \right)^2$	$n_{..}$
Between Herds	$n_{..}$	$\sum_1 \left(\frac{\sum_k \frac{n_{ijk}}{n_{1.k}}}{n_{1.}}$	Number of Herds
Correction Term	$\frac{\sum_1 n_{1.}^2}{n_{..}}$	$\frac{\sum_{ij} \left(\frac{\sum_k n_{ijk}}{n_{1.k}} \right)^2}{n_{..}}$	1

necessary to build into the models which describe average production of dams and daughters the additive effects of their genes. The model for the average production of the dams then becomes,

$$\frac{D_{1.k}}{n_{1.k}} = \mu + h_1 + \sum_j \frac{n_{ijk}}{n_{1.k}} y_{1j} + g_{1k} + \sum_j \frac{e_{ijk}}{n_{1.k}}$$

with the definitions of all subscripts remaining the same as in the original model. g_{1k} is defined as the effect of the genotype

of the k^{th} dam in the i^{th} herd. It includes all the additive effects plus a small portion of the epistatic effects. e_{ijk} includes all permanent and temporary environmental effects plus all the effects of dominance and most of the epistatic effects.

The model describing the average production of the daughters is,

$$\frac{O_{i.kl}}{n_{i.kl}} = \mu + h_i + \sum_j \frac{n_{1jkl} y_{1j}}{n_{i.kl}} + \frac{S_{1k}}{2} + \sum_j \frac{e_{1jkl}}{n_{i.kl}}$$

$O_{i.kl}$ refers to the total production of the l^{th} daughter of the k^{th} dam in the i^{th} herd. μ in this case refers to the general mean of the population of daughters. It is not necessary to assume that μ for the dams and μ for the daughters have the same value, since this component of covariance goes out with the correction term. All other letters including subscripts have the same meaning as in the model given for the average production of the dams.

Since a daughter receives a sample half of the genes which her dam had, the effect of the transmissible part of the dam's additively genetic effects will on the average be only one-half as great on the daughter's records as on the records of the dam.

The theoretical composition of the crossproducts in the uncorrected total and between herds and in the correction term is given in table 8. Note that no error covariance component appears. For this to be true the permanent and temporary environmental effects must not be correlated between daughter and dam. The sampling nature of inheritance gives assurance that

the dominance deviations and the bulk of the epistatic deviations are not correlated between parent and offspring. The intra-herd environmental effects on dam and daughter would not be correlated unless they were highly contemporary and the year effects had not been removed, or breeders tried intentionally to treat daughter and dam alike but differently from other pairs in those intra-herd variations in treatment which were under their control.

Table 8. Theoretical Composition of Uncorrected Crossproducts and Correction Term. Average Fat Production of Daughters and Dams.

Source of Variation	H	y^a	$\frac{G}{2}$
Total	$n...$	$\sum_{ijkl} \frac{n_{ijk} n_{ijkl}}{n_{i.k} n_{i.kl}}$	$n...$
Between Herds	$n...$	$\sum_i \frac{\sum_j \left(\sum_{k_a} \frac{n_{ijk_a}}{n_{i.k}} \sum_{kl} \frac{n_{ijkl}}{n_{i.kl}} \right)}{n_{i..}}$	$\sum_i \frac{\sum_k n_{ik.}^2}{n_{i..}}$
Correction Term	$\frac{\sum_i n_{i..}^2}{n...}$	$\sum_{ij} \frac{\left(\sum_{k_a} \frac{n_{ijk_a}}{n_{i.k}} \sum_{kl} \frac{n_{ijkl}}{n_{i.kl}} \right)}{n...}$	$\frac{\sum_{ik} n_{ik.}^2}{n...}$

^a k_a refers to the a^{th} comparison for the k^{th} dam.

Since some of the dams have more than one daughter, the number of $G/2$ components which go out with the herd crossproduct is greater than the number of herds. The number of $G/2$

components in the corrected crossproduct within herds is,

$$n_{...} - \sum_1 \frac{\sum_k n_{ik.}^2}{n_{1..}}$$

and will always be less than $n_{...}$ minus the number of herds if $n_{ik.}$ is ever greater than 1. The expected number of $G/2$ components in the various uncorrected crossproducts is based simply on the biological fact that the expected resemblance between a dam and any one of her daughters is the same. In algebraic symbols, the assumption is made that

$$\frac{D_{1.k} \times O_{1.kl}}{n_{1.k} n_{1.kl}}$$

has the same expected value for each of the 1 daughters where the other subscripts remain the same.

The analysis of the variances and covariances necessary to estimate the heritability of type and production and the genetic and phenotypic correlations between type and production from the 2,786 daughter-dam pairs is given in tables 9 and 10. T and P stand for type and production and D and O stand for dam and offspring respectively. The theoretical compositions of all mean squares and covariances were determined by the same method as described above for P_D and P_DP_O .

1. Heritability of type.

The regression of type of daughter on type of dam within herds is,

$$\frac{\frac{G_T}{2}}{E_{TD}} = \frac{.03898}{.54950} = .071$$

Heritability of differences in type within herd is therefore estimated to be $.142 \pm .042$, i.e., twice the regression of daughter on dam, from these data. This estimate is only half that given by Tyler and Hyatt (1948) for Ayrshires from an intra-sire regression of daughter on dam (1,601 pairs) where all daughters and dams within sire were classified by the same judge on the same day. Most of the differences between judges are eliminated from the intra-herd regression given here, since most cows in a given herd were classified by one judge and on the same day. Since this was not always the case, some judge differences remain in both $G_T/2$ and E_{TD} . The number of judge components which would be expected in $G_T/2$ would be,

$$\frac{\sum_{ijkl} n_{ijk} n_{ijkl} - \sum_i \frac{\sum_j (n_{ij.} n_{ij..})}{n_{i...}}}{2432.2}$$

Where the i^{th} daughter of the k^{th} dam was classified by the j^{th} judge in the i^{th} herd. $n_{ij.}$ is summed over all pairs and not just over all k as the dot might indicate. The number of judge components expected to be present in E_{TD} is,

$$\frac{n_{...} - \sum_i \frac{\sum_j n_{ij.}^2}{n_{i..}}}{1818}$$

$\sum_{ijkl} n_{ijk} n_{ijkl}$ will equal $n_{....}$ if both daughter and dam are classified by the same judge in all pairs; otherwise, it will always be less than the total number of daughter-dam pairs.

Table 9. Analysis of Variances of Type and Average Fat Production of Daughters and Dams

Source of Variation	d.f.	Sum of Sqs.	Mean Sq.	Theoretical Composition of Mean Square
Total	2043	$\frac{T_D}{1,359.1}$	$\frac{T_D}{.66525}$	$E_{TD} + .992 H_T$
Between Herds	225	360.1	1.60044	$E_{TD} + 9.012 H_T$
Within Herd	1818	990.0	.54950	E_{TD}
<hr/>				
Total	2785	$\frac{T_O}{1,697.4}$	$\frac{T_O}{.60948}$	$E_{TO} + .992 H_T$
Between Herds	225	408.7	1.81644	$E_{TO} + 12.280 H_T$
Within Herd	2560	1,288.7	.50340	E_{TO}
<hr/>				
Total	2043	$\frac{P_D}{18,772,618}$	$\frac{P_D}{9,189}$	$E_{PD} + .531 Y + .992 H_P$
Between Herds	225	9,358,618	41,594	$E_{PD} + 2.717 Y + 9.012 H_P$
Within Herd	1818	9,414,000	5,178	$E_{PD} + .260 Y$
<hr/>				
Total	2785	$\frac{P_O}{25,789,507}$	$\frac{P_O}{9,260}$	$E_{PO} + .684 Y + .992 H_P$
Between Herds	225	13,673,116	60,769	$E_{PO} + 2.857 Y + 12.280 H_P$
Within Herd	2560	12,116,391	4,733	$E_{PO} + .493 Y$

Summary of Components

	$\frac{E_T}{.5495}$	$\frac{E_P}{5044}$	$\frac{H_T}{.1166}$	$\frac{H_P}{8900}$
Dams				
Daughters	.5034	4479	.1069	4464

Table 10. Analysis of Covariances of Type and Average Fat Production of Daughters and Dams

Source of Variation	d.f.	Crossproduct	Cov.	Theoretical Composition of Covariance
		$T_D T_O$	$T_D T_O$	
Total	2784.3	386.7	.13889	$G_T/2 + .992 H_T$
Between Herds	351.4	291.9	.83077	$G_T/2 + 7.864 H_T$
Within Herd	2432.2	94.8	.03898	$G_T/2$
		$P_D P_O$	$P_D P_O$	
Total	2784.3	13,397,819	4,812	$G_P/2 + .204 Y + .992 H_P$
Between Herds	351.4	11,900,107	33,869	$G_P/2 + 1.532 Y + 7.864 H_P$
Within Herd	2432.2	1,497,712	616	$G_P/2 + .012 Y$
		$T_D P_O$	$T_D P_O$	
Total	2784.3	11,559	4.152	$Cov(G_T G_P)/2 + .992 Cov_H(T_D P_O)$
Between Herds	351.4	9,745	27.736	$Cov(G_T G_P)/2 + 7.864 Cov_H(T_D P_O)$
Within Herd	2432.2	1,814	.746	$Cov(G_T G_P)/2$
		$P_D T_O$	$P_D T_O$	
Total	2784.3	12,891	4.630	$Cov(G_T G_P)/2 + .992 Cov_H(P_D T_O)$
Between Herds	351.4	10,328	29.394	$Cov(G_T G_P)/2 + 7.864 Cov_H(P_D T_O)$
Within Herd	2432.2	2,563	1.054	$Cov(G_T G_P)/2$
		$T_D P_D$	$T_D P_D$	
Total	2043	21,852	10.696	$Cov(TP) + .992 Cov_H(TP)$
Between Herds	225	7,014	31.171	$Cov(TP) + 9.012 Cov_H(TP)$
Within Herd	1818	14,838	8.162	$Cov(TP)$
		$T_O P_O$	$T_O P_O$	
Total	2785	35,119	12.610	$Cov(TP) + .992 Cov_H(TP)$
Between Herds	225	16,954	75.352	$Cov(TP) + 12.280 Cov_H(TP)$
Within Herd	2560	18,165	7.096	$Cov(TP)$

Summary of Components

$$G_T/2 = .0390$$

$$H_T = .1007$$

$$G_P/2 = 609.7$$

$$H_P = 4129$$

$$Cov(G_T G_P)/2 = .900 \text{ (Average)}$$

$$Cov_H(T_D P_O) = 3.432$$

$$Cov_H(P_D T_O) = 3.604$$

	$Cov(TP)$	$Cov_H(TP)$
Dams	8.162	2.553
Daughters	7.096	5.558

It would be zero if one judge classified all the dams within herd and another judge classified all the daughters, in which case $\sum_i \frac{\sum_j (n_{1j} \cdot n_{1j..})}{n_{1...}}$ would also be zero.

In this case

$$\sum_i \frac{\sum_j n_{1j}^2}{n_{1..}}$$

would equal $n_{1..}$ and judge

differences, even though real, would not bias the regression of daughter on dam.

In those herds classified more than once some of the daughters and dams will have been classified by the same judge but in other pairs one judge will have classified the daughters and a different judge at a different time will have classified the dams. The expected distribution of judge components in $G_T/2$ and E_{TD} would then be very similar to that of the Y component for production in $G_P/2$ and in E_{PD} , respectively. In other words, the judge differences will have increased the variance of the dams (E_{TD}) about 10 to 20 times as much as they increased the covariance between dam and daughter ($G_T/2$). Consequently, the estimate of heritability for type given here is expected to be too low if judge differences are real. Perhaps the difference between the intra-herd regression found here and the intra-sire and intra-judge regression reported by Tyler and Hyatt is some indication that judge differences are important. On the other

hand, the heritability of differences in type might actually be lower in Jerseys than in Ayrshires. A third possibility is that the two values are equally valid and unbiased estimates of the same parameter. This interpretation cannot be overlooked because the sampling errors are so large that a difference this large has a fairly high probability of occurring by chance.

As seen in table 9 the dams were more variable in type ratings than their daughters. If breeders were allowing cows to become dams when they had better than average type, regardless of their production, and were at the same time retaining all cows above a certain level of production with practically no regard to their type rating, the variance of the dams would be only slightly different from that of all cows. Table 6 pointed out that this was the type of selection actually practiced. The culling levels would be expected to shift considerably from one herd to another and from time to time within the same herd. Actually the dams and all cows (including all dams) were about equally variable within herd -- .5495 for dams and .5430 for all cows. A closer genetic relationship among the daughters would be expected than among the dams, since 554 dams had two or more daughters each. This would have the effect of reducing the variance in the daughters, but it seems unlikely that it could account for all the difference between the intra-herd variances of dams and daughters observed here. Perhaps a more important reason for this is the regulation of the A.J.C.C. which prohibits a cow from being classified "Excellent" until she has

dropped her second calf. Many more daughters than dams would not have had a second calf at the time of classification. Possibly also the judges are more conservative when classifying young cows, knowing that the rating they give can never be lowered but may be raised on reclassification, according to the regulations of the classification program.

Three separate estimates of the herd component for type are available from tables 9 and 10. These are presented in table 11, along with the percentage of the total variance (.6590)

Table 11. Estimates of Herd Differences in Type

Source of Estimate	H_T	5 Per Cent Fiducial Limits	Percentage of Total Variance All Cows
Dams	.1166	.0905--.1641	17.7
Daughters	.1069	.0847--.1448	16.2
Covariance	.1007	.0841--.1255	15.3

for all cows that each estimate represents. Differential selection with respect to type from one herd to another could account for the differences between herds being about 1.5 per cent greater for the dams than daughters. However, the approximate 5 per cent fiducial limits on H_T for dams and daughters¹ overlap, indicating that the two estimates could very well be estimates of the same parameter.

¹These fiducial limits were established by the method outlined by Satterthwaite (1946). He points out that, even though the distribution used is approximate, it approaches the exact distribution when the number of degrees of freedom is large.

2. Heritability of production

From an analysis of all fat production records (including those made by cows without an official type rating) as outlined in section IV, A, 3 and carried out by Legates (1949), the Y component was estimated to be 515.5. This represented 4.9 per cent of the total variance. There appears to be no logical reason why the year effects within herd should be different for those dams or their daughters which had a type rating than for those which did not. Therefore, 515.5 was used for Y when computing the other components in tables 9 and 10.

The intra-herd regression of daughter's production on dam's production, presumably freed from year differences, is

$$\frac{\frac{G_P}{2}}{E_{PD}} = \frac{609.7}{5,044} = .121$$

From this the heritability of differences in fat production records made in the same herd and year, when adjusted to a single-record basis by the formula given by Lush and Straus (1942), is estimated to be $.177 \pm .030$. Ignoring the effects of years, the intra-herd regression of daughter on dam is .119, and heritability on a single-record basis is $.167 \pm .029$. The repeatability figures used in adjusting these regressions to a single-record basis were .46 and .41 respectively. These are the estimates obtained by Legates (1949) from all 12,405 cows.

There are two primary reasons for the heritability of fat production being only slightly higher with the effects of years

removed than when these year effects were ignored. First, by using an average of all records available on the dam the amount of the year component remaining in the intra-herd variance of the dams is diminished considerably from what it would be if only one record was used. Second, the differences between years within herd accounted for only a small portion of the intra-herd variance.

Unless there is a deliberate effort to analyze only those records of a dam and her daughter that were made in the same year, the year effects in the intra-herd covariance will always be small. If the first record of dam and of daughter are used there will be no year component at all in the covariance.

The production of the dams within herd was just as variable as the production of all cows (5,178 for dams and 5,171 for all cows), even though the dams were a select group. The dams averaged only about 5 pounds of butterfat more than all cows, however, indicating that selection on fat production was not very intense. This, combined with the simultaneous selection for type, leads one to expect little, if any, decrease in the variance of the dams. The reasons for the significantly lower variation in the daughters' production are not clear. It can not be explained by the type of selection practiced on the dams. A part of it could be due to the closer intra-herd relationship of the daughters as compared to the dams, which was pointed out previously when discussing the same situation with type. On the other hand, the dams had more records each than the

daughters and this should tend to make the dams less variable.

The three separate estimates of the herd component for production and the percentage of the total variance (9,085) for all cows that each represents when year effects are ignored are given in table 12. When the year component is considered, i.e., if differences between herds are calculated on the intra-year basis, each of the estimates given in table 12 decreases

Table 12. Estimates of Herd Differences in Fat Production Ignoring the Year Effects

Source of Estimate	H _p	5 Per Cent Fiducial Limits	Percentage of Total Variance All Cows
Dams	4041	3324--5151	44.5
Daughters	4563	3788--5737	50.2
Covariance	4229	3547--5234	46.5

approximately one hundred or about 2.4 per cent. Although there is some indication that differences between herds were greater for daughters than dams, the 5 per cent fiducial limits show that this difference is not at all significant.

3. Genetic correlation

The genetic correlation between type and production was estimated by using the four intra-herd genetic covariances given in table 10, as follows,

$$\frac{\sqrt{(1.054)(.746)}}{\sqrt{(616)(.03898)}} = \frac{.8865}{4.900} = .181$$

The effects of differences between years in the same herd are present in only the one covariance, $G_p/2$, and even then in a very limited amount. Since the expected value of either one of the four covariances is the same, regardless of the number of records per cow, the estimate of the genetic correlation is automatically on a single-record basis.

As pointed out by Hazel (1943), the sampling error of a genetic correlation is necessarily very large, since it is calculated from four statistics which have some correlation between their sampling errors. A method of obtaining fiducial limits on estimates of genetic correlations has not been worked out.

4. Phenotypic correlations

The 6 possible intra-herd phenotypic correlations involving type and production of daughter and dam were calculated (1) by eliminating the effects of years on the average production, and (2) by ignoring the year effects. All correlations were then adjusted to a single-record basis, using .46 for repeatability of records when the year effects were removed and .41 when ignoring years. These correlations are presented in table 13.

The correlations between type and production in the same animal (.153 for dams and .145 for daughters, or .123 and .124 when reduced to the single-record basis) agree very closely with the one obtained by Tyler and Hyatt (1948) in a study with 5,177 Ayrshire cows. They had an average of 3.5 records per cow

Table 13. Intra-Herd Phenotypic Correlations Involving Type and Production of Daughter and Dam.

	T_D	P_D	P_D
T_O	Excluding Year Effects	----	.1494
	Single Record	----	.1290
	Ignoring Year Effects	.0741	.1454
	Single Record	.0741	.1235
T_D	Excluding Year Effects		.0150
	Single Record		.0130
	Ignoring Year Effects		.0146
	Single Record		.0124
P_O	Excluding Year Effects		.1283
	Single Record		.0913
	Ignoring Year Effects		.1244
	Single Record		.0852

and obtained an intra-herd correlation of .19 between type rating and average production. When adjusted to a single record basis, this becomes .144, assuming that repeatability of their fat production records was .4.

C. Paternal Sister Analysis

The linear mathematical model used to describe the average production of a given cow was,

$$\frac{P_{i.kl}}{n_{i.kl}} = \mu + h_i + \sum_j \frac{n_{ijk1} y_{ij}}{n_{i.kl}} + s_{il} + e_{ikl}$$

Where:

$P_{i.kl}$ = sum of all the available records of the k^{th} daughter of the l^{th} sire in the i^{th} herd.

$n_{i.kl}$ = number of records for the k^{th} daughter.

μ = general mean, characteristic of all records.

h_i = effect of the i^{th} herd.

n_{ijk1} = number of records for the k^{th} daughter of the l^{th} sire in the j^{th} year and i^{th} herd.
Usually either 1 or 0.

y_{ij} = effect of the j^{th} year in the i^{th} herd.

$\sum_j \frac{n_{ijk1} y_{ij}}{n_{i.kl}}$ = yearly effect remaining in the average of all records of that cow.

s_{il} = effect of the l^{th} sire in the i^{th} herd.

$$e_{ikl} = p_{ikl} + \sum_j \frac{e_{ijk1}}{n_{i.kl}}$$

and

p_{ikl} = effect of the k^{th} daughter of the l^{th} sire in the i^{th} herd.

e_{ijk1} = random error associated with each record.
Assumed $N(0, \sigma_e^2)$.

The only difference between this model and the one used for the average production of the dams (Page 40) is the additional term, s_{il} , to describe the effect of the sire within herd on

the average production of the cow.

The theoretical composition of the sums of squares in the uncorrected total, between sires, and between herds and in the correction term is given in table 14 for fat production. A comparison of the number of sire components in the various sums of

Table 14. Theoretical Composition of Uncorrected Sums of Squares and Correction Term. Average Fat Production of All Cows.

Source of Variation	H_P	Y	S_P	E_P
Total	$n_{...}$	$\sum_{ijkl} \frac{n_{ijkl}^2}{n_{i.kl}}$	$n_{...}$	$n_{...}$
Between Sires	$n_{...}$	$\sum_{il} \frac{\sum_j \sum_k \frac{n_{ijkl}^2}{n_{i.kl}}}{n_{i..}}$	$n_{...}$	Number of Sires
Between Herds	$n_{...}$	$\sum_i \frac{\sum_j \sum_k \frac{n_{ijkl}^2}{n_{i.kl}}}{n_{i..}}$	$\sum_i \frac{n_{i..}^2}{n_{i..}}$	Number of Herds
Correction Term	$\frac{\sum_i n_{i..}}{n_{...}}$	$\frac{\sum_{ij} \sum_k \frac{n_{ijkl}^2}{n_{i.kl}}}{n_{...}}$	$\frac{\sum_{il} n_{i..}^2}{n_{...}}$	1

squares and mean squares with a similar model for proportionate and disproportionate frequencies is given in Appendix B.

In order for S_P (or S_T for type) to contain 1/4 of the additively genetic variance it is necessary that the sires

within a herd be unrelated and have been mated to a random sample of unrelated dams in that herd. If the sires are related the p_{ikl} will be correlated between sires within herd and S will contain less than 1/4 of the additively genetic variance, but if the dams to which each sire is mated are related the p_{ikl} will be correlated within sire and S will contain more than 1/4 of the genic variance. S may also contain an environmental component if the e_{ijkl} are correlated for daughters of the same sire.

The model used to describe the type rating of a cow was,

$$T_{ijk} = \mu + h_i + s_{ij} + e_{ijk}$$

Where:

T_{ijk} = type rating of the k^{th} daughter of the j^{th} sire in the i^{th} herd.

μ = general mean for type

h_i = effect of the i^{th} herd.

s_{ij} = effect of the j^{th} sire in the i^{th} herd.

e_{ijk} = random error associated with each type rating plus the cow effect.

Since each cow used in this analysis had a type rating and one or more production records, the number of H_T , S_T , and E_T components in the uncorrected sums of squares and in the correction term are expected to be the same as calculated for H_P , S_P , and E_P respectively.

Because the date of classification or the judge who did the classifying were not listed in our material, the effects of

years or judges could not be investigated. The model for type assumes that judge and year differences are zero or that such effects are distributed at random among the daughters of a sire; i.e., that any effects of these differences are in e_{ijk} . Since most of the daughters of a sire would be classified on the same day by the same judge, S_T will be an overestimate of the genetic differences between sires if judge or year differences are real.

The models used and the procedure followed in obtaining the expected number of the covariance components in the uncorrected crossproducts and in the correction term were given in section IV, A, 2.

The 8,464 cows which had a type rating and one or more fat production records were actually sired by 2,605 different sires. However, in order to obtain an estimate of S_P , S_T , and $Cov_s(TP)$ on an intra-herd basis it was necessary to consider a sire separately for each herd in which he had a daughter. Since there were 324 sires with daughters in two or more herds, the total number of sires for calculating degrees of freedom was 3,042.

The analyses of variance and covariance of type and average fat production together with the theoretical composition of the mean squares and covariances are presented in table 15. This analysis affords no estimate of Y . Its value was assumed to be 515.5, as obtained by Legates from all 12,405 cows by the method discussed in section IV, A, 3. The distribution of the records of production of classified cows within herd should be

very nearly the same as for all cows. There appears to be no logical reason why the records of classified cows should vary more or less from one year to another within herd than do the records of all cows. It seems likely therefore, that little error in the estimates of other components for fat production would result from assuming that Y is 515.5 in these data.

The estimates of all components of variance and covariance for type and production from the paternal sister analysis are given in table 16. The correlations between paternal half sisters for type, average production, for type in one and production in another, and between the genotype of the sire for type and his genotype for production are given at the bottom of the table.

The average number of records per cow was 2.01. Therefore, r_{pp} is $\frac{2.01}{1+(1.01).46}$ times as large as would be expected on a single-record basis. The reciprocal of this times the .166, yields .121 as what r_{pp} would have been if only one record per cow had been used. If one could assume that this represents only the correlation between the genic effects which are common to paternal sisters and does not include any environmental correlation, then the heritability of differences in single records of fat production within herd would be estimated to be .484. The assumption that $p = 0$ seems invalid, considering the estimate of .167 for heritability of single records from the daughter-dam analysis where there would be much less reason to suspect an environmental correlation.

Table 15. Analysis of Variance and Covariance of Type and Average Fat Production

Source of Variation	d.f.	Sum of Sqs. or Crossproduct	Mean Sq. or Cov.	Theoretical Composition of Mean Sq. or Cov.
Total	8463	\underline{P} 76,888,558	\underline{P} 9,085	$E_p + .647 Y + .999 S_p + .994 H_p$
Between Sires	3041	53,002,650	17,429	$E_p + 1.241 Y + 2.780 S_p + 2.765 H_p$
Between Herds	244	34,385,924	140,926	$E_p + 8.736 Y + 6.050 S_p + 34.466 H_p$
Within Herd	8219	42,502,634	5,171	$E_p + .407 Y + .849 S_p$
Between Sires	2797	18,616,726	6,656	$E_p + .587 Y + 2.495 S_p$
Within Sire	5422	23,885,908	4,405	$E_p + .314 Y$
Total	8463	\underline{T} 5,577.1	\underline{T} .65900	$E_t + .999 S_t + .994 H_t$
Between Sires	3041	2,987.7	.98247	$E_t + 2.780 S_t + 2.765 H_t$
Between Herds	244	1,114.3	4,56680	$E_t + 6.050 S_t + 34.466 H_t$
Within Herd	8219	4,462.8	.54299	$E_t + .849 S_t$
Between Sires	2797	1,873.4	.66979	$E_t + 2.495 S_t$
Within Sire	5422	2,589.4	.47757	E_t
Total	8463	\underline{TP} 97,351.7	\underline{TP} 11.5032	$Cov_E(TP) + .999 Cov_S(TP) + .994 Cov_H(TP)$
Between Sires	3041	61,552.1	20.2407	$Cov_E(TP) + 2.780 Cov_S(TP) + 2.765 Cov_H(TP)$
Between Herds	244	36,377.1	149.0365	$Cov_E(TP) + 6.050 Cov_S(TP) + 34.466 Cov_H(TP)$
Within Herd	8219	60,974.6	7.4187	$Cov_E(TP) + .849 Cov_S(TP)$
Between Sires	2797	25,175.0	9.0007	$Cov_E(TP) + 2.495 Cov_S(TP)$
Within Sire	5422	35,799.6	6.6027	$Cov_E(TP)$

Eliminating the yearly effects within herd decreased r_{PP} only .004 -- from .170 to .166. If the number of year components were the same in the mean squares between sires within herd and also within sires, the elimination of yearly effects would not change S_P but would increase the correlation between paternal half sisters since the denominator would be decreased.

Table 16. Estimates of Components of Variance and Covariance

Variance Components		Covariance Components
T	P	TP
$E_T = .47757$	$E_P = 4243$	$Cov_E(TP) = 6.6027$
$S_T = .07703$	$S_P = 846$	$Cov_S(TP) = .9610$
$H_T = .10512$	$H_P = 3687$	$Cov_H(TP) = 3.9654$
$r_{TT} = \frac{S_T}{E_T + S_T}$	$r_{PP} = \frac{S_P}{E_P + S_P}$	$x = \frac{Cov_S(TP)}{\sqrt{S_T S_P}}$
$= .139$	$= .166$	$= .119$
$r_{TP} = r_{PT} = \frac{Cov_S(TP)}{\sqrt{E_T E_P}} = .021$		

The daughters of a sire have most of their records in one or two consecutive years in the present data. Consequently almost twice as much of the year component appears in the mean square between sires within herd as in the mean square within sires. If one record per daughter had been selected so that every daughter of a given sire had her record in the same year, no year components

would have been present in the mean square within sires and almost one year component would have been expected in the mean square between sires within herds. By ignoring year effects in such a case the half sister correlation would be considerably overestimated. By using averages of all records available on each cow so much of this bias was eliminated that the additional labor required to eliminate the remaining yearly effects was not justified.

The intra-herd phenotypic correlation between type and average fat production in the same individual was

$$\frac{7.4187}{\sqrt{(4961.5)(.54299)}} = .143$$

for all 8,464 cows when adjusted for yearly differences in production. When adjusted to a single-record basis for type and production this was .122. Ignoring yearly differences within herd on fat production r_{TP} was .140 and adjusted to a single-record basis it was .117. The phenotypic correlation between type and production between herd averages was .186 ignoring yearly effects. Herds that averaged one classification grade higher also averaged about 33 more pounds of butterfat per cow. Cows that averaged one grade higher in type than other cows in the same herd averaged only about 14 more pounds of butterfat per cow.

Many of the mates of a sire within herd are likely to be related. Lush and Straus (1942) have estimated that the average relationship among the mates of a sire within herd is about

.10. This increases the fraction of the genes which the paternal sisters are expected to have alike--from .25 to about .275.

Assuming that the average relationship among paternal sisters (r) is .275, heritability of type (g_T^2) is .142, and heritability of single records of fat production (g_P^2) is .177, then r_{TT} and r_{PP} (on a single-record basis) are each only about .004 larger than if all mates of a sire were unrelated.

As pointed out in section IV, A, 2, by using the estimates of g_T^2 , g_P^2 , and \underline{x} as obtained from the daughter-dam analysis one can estimate the environmental correlations (\underline{w} , \underline{t} , \underline{p} , and \underline{u}) and the environmental path coefficients (e_T and e_P) for paternal sisters. These genetic and environmental relations presumed to underly the observed phenotypic correlations adjusted to a single-record basis in paternal sisters for type and fat production are presented in their respective positions in the path coefficient diagram shown in figure 3.

Even though \underline{x} is larger than \underline{w} (.182 as compared to .111) it accounts for only about 24 per cent of the phenotypic correlation between type and production. The reason for this is apparent from the genetic and environmental path coefficients, i.e., both traits are largely determined by the environment. Since environment that is conducive to high fat production would, from a priori grounds, be expected to improve the general health and appearance of a cow and vice versa, it is not surprising that \underline{w} is as high as .111. The estimates of \underline{t} and \underline{p} do appear a little higher than one would expect. Under the assumptions

made they account for 72 and 60 per cent of the observed phenotypic correlations respectively.

Most of the daughters of a given sire are likely to have been classified by the same judge on the same day. This makes $r_{TT'}$ and t too large if judge or year differences are real.

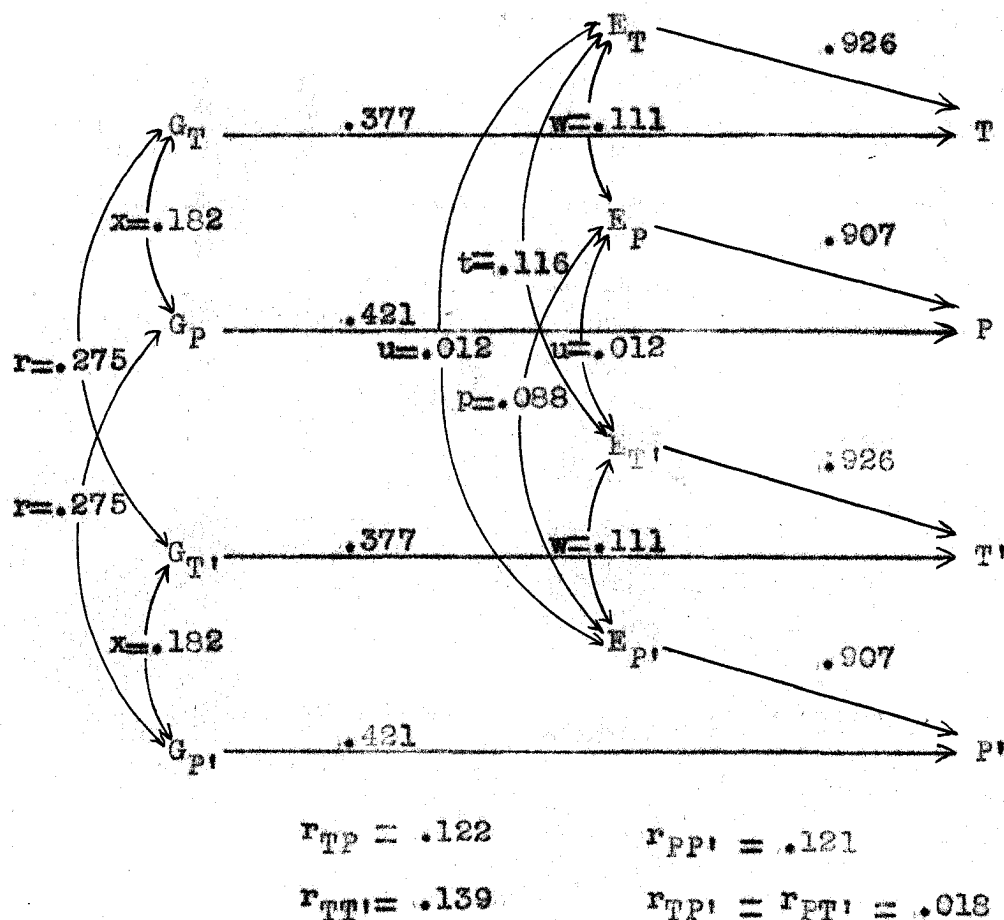


Figure 3. Path Coefficient Diagram to Illustrate the Genetic and Environmental Relations Underlying Phenotypic Correlations Among Paternal Sisters.

Perhaps this explains why $r_{TT'}$ is larger than $r_{PP'}$ even though heritability is presumably slightly lower for type, rather than the alternate explanation of more of the permanent environmental effects in common among paternal sisters for type than production.

Although the environmental correlations in figure 3 seem to be major causes of the observed phenotypic correlations or covariances they should have practically no effect on the estimate of the genetic correlation from paternal sisters, since

$$\sqrt{\frac{u}{t \ p}} = .122$$

In other words, the contribution of the environmental effects to the numerator and denominator of the genetic correlation fraction is in about the same ratio as the genetic correlation obtained in the daughter-dam analysis where these environmental correlations were not likely to have been present.

D. Maternal Sister Analysis

The dam of each of the 8,464 cows used in the paternal sister analysis was known, so that by classifying them by dam instead of sire the relations among maternal sisters were studied. The models and the procedure used in determining the theoretical composition of the mean squares and covariances underlying this analysis were essentially the same as those given for the paternal sister analysis. The only change necessary is a substitution of the symbol for the dam effects in place of that used for the sire effects.

The 8,464 cows were out of 6,912 dams when a dam with a daughter in more than one herd was counted as a different dam for each herd in which she had a daughter. This procedure was necessary in order to obtain an estimate of the dam component (D) on a strictly intra-herd basis. Approximately 50 dams had daughters in more than one herd. The frequency distribution of the number of daughters per dam and the average type and fat production of the daughters in each class are given in table 17. The average production of the dams increases steadily with increasing number of daughters per dam, at least up to the group with five daughters whose numbers are so small that their average is not dependable. The average type on the other hand shows no such consistent trend. There are two possible explanations for production increasing with the number of daughters: (1) a larger number of dams in the higher producing herds than

in the lower ones may have had two or more daughters, or (2) breeders were selecting for rebreeding those dams which already had daughters with higher than average production records. It is doubtful that the second explanation could account for much of the difference observed from the fact that after a daughter is born the cow is rebred about two and one-half years before that daughter can possibly show whether she is a high producer or not. Therefore most of the second or third daughters would already have been born before ever the breeder knew what the first daughter's production was.

Table 17. Frequency of Number of Daughters Per Dam and Average Fat Production and Type of Daughters in Each Class.

No. Daughters Per Dam	No. Dams	Average Production	Average Type
1	5703	427.7	4.19
2	936	437.1	4.17
3	214	456.8	4.21
4	48	478.6	4.14
5	11	447.5	4.47
Totals	6912	433.3	4.19

The analyses of variance and covariance of type and fat production plus the expected composition of the mean squares and covariances for maternal sisters are given in table 18. The sums of squares and crossproducts in the total, between herds, and within herds are identical with those given in table 15 for the paternal sister analysis since all cows with type were used in both analyses.

Table 18. Analysis of Variance and Covariance of Type and Average Fat Production

Source of Variation	d.f.	Sum of Sqs. or Crossproduct	Mean Sq. or Cov.	Theoretical Composition of Mean Sq. or Cov.
<u>P</u>				
Total	8463	76,888,558	9,085	$E_P + .647 Y + 1.000 D_P + .994 H_P$
Between Dams	6911	69,864,514	10,109	$E_P + .716 Y + 1.224 D_P + 2.765 H_P$
Between Herds	244	34,385,924	140,926	$E_P + 3.736 Y + 1.416 D_P + 34.466 H_P$
Within Herd	8219	42,502,634	5,171	$E_P + .407 Y + .988 D_P$
Between Dams	6667	35,478,590	5,322	$E_P + .423 Y + 1.218 D_P$
Within Dam	1552	7,024,044	4,526	$E_P + .339 Y$
<u>T</u>				
Total	8463	5,577.1	.65900	$E_T + 1.000 D_T + .994 H_T$
Between Dams	6911	4,847.1	.70136	$E_T + 1.224 D_T + 2.765 H_T$
Between Herds	244	1,114.3	4.56680	$E_T + 1.416 D_T + 34.466 H_T$
Within Herd	8219	4,462.8	.54299	$E_T + .988 D_T$
Between Dams	6667	3,732.8	.55989	$E_T + 1.218 D_T$
Within Dam	1552	730.0	.47036	E_T
<u>TP</u>				
Total	8463	97,351.7	11.5032	$Cov_E(TP) + 1.000 Cov_D(TP) + .994 Cov_H(TP)$
Between Dams	6911	87,981.5	12.7306	$Cov_E(TP) + 1.224 Cov_D(TP) + 2.765 Cov_H(TP)$
Between Herds	244	36,377.1	149.0865	$Cov_E(TP) + 1.416 Cov_D(TP) + 34.466 Cov_H(TP)$
Within Herd	8219	60,974.6	7.4187	$Cov_E(TP) + .988 Cov_D(TP)$
Between Dams	6667	51,604.4	7.7403	$Cov_E(TP) + 1.218 Cov_D(TP)$
Within Dam	1552	9,370.2	6.0375	$Cov_E(TP)$

The components of variance and covariance, as obtained from the maternal sister analysis in table 18, are given in table 19. The correlations between maternal sisters in type and average production and the genetic correlation are given at the bottom of the table. $r_{PP'}$ and $r_{TP'}$ when adjusted to a single-record basis reduce to .091 and .026 respectively.

Table 19. Estimates of Components of Variance and Covariance

Variance Components		Covariance Components
T	P	TP
$E_T = .47036$	$E_P = 4351$	$Cov_E(TP) = 6.0375$
$D_T = .07354$	$D_P = 618$	$Cov_D(TP) = 1.3986$
$H_T = .10022$	$H_P = 3807$	$Cov_H(TP) = 4.0930$
$r_{TT'} = \frac{D_T}{E_T + D_T}$	$r_{PP'} = \frac{D_P}{E_P + D_P}$	$r = \frac{Cov_D(TP)}{\sqrt{D_T D_P}}$
$= .135$	$= .124$	$= .208$
$r_{TP'} = r_{PT'} = \frac{Cov_D(TP)}{\sqrt{E_T E_P}} = .031$		

The average number of dams per herd was about 28, whereas the average number of sires per herd was only 12. Consequently, many of the dams mated to the same sire within herds were paternal sisters and one full E component is not actually present in the mean squares between dams within herds. No attention was paid to the presence of such comparisons in calculating D_T and

D_p as given in table 19; hence they are slightly underestimated. If $1/6$ of the comparisons between dams within herd were between paternal half sisters the mean square for production between dams within herds would be approximately $1/6 \times .166$ or 2.76 per cent too small, and the corresponding mean square for type about $1/6 \times .139$ or 2.31 per cent too small. The increase in D_p and D_T and also r_{pp} and r_{TT} would be $\frac{1}{1.218}$ of these percentages, since any increase in the mean square would be divided by the coefficient for the dam component.

It will also be noted that E_p for maternal sisters is about 108 greater than E_p for paternal sisters. A possible explanation for this lies in the fact that the higher producing herds had higher intra-herd variance, as pointed out in section IV, and that more of the daughters out of dams which had more than one daughter probably were in such higher producing herds. On the other hand, there were 1,108 full sisters among the 2,761 daughters out of dams with a multiple number of daughters from which E_p in table 19 is estimated. These full sisters make both E_p and E_T too small. The effect of the full sisters on these two components can be eliminated by subtracting the sum of squares and degrees of freedom between full sisters (within sets of full sisters) from the corresponding sum of squares listed as within dams in table 18. The year components are subtracted out in the same manner. This adjustment is presented in table 20 together with the estimates of E_p , E_T , D_p , D_T , r_{pp} , and r_{TT} .

which are all presumably freed thus from the effect of full sisters, so that table 20 presents the evidence from maternal half sisters only. Adjusted to a single-record basis $r_{pp1} = .080$.

Table 20. Removal of Full Sister Effect on E_p and E_T of Maternal Sister Analysis

Source of Variation	d.f.	Sum of Squares		Mean Squares	
		P	T	P	T
Between 1/2 and Full Sisters ^a	1552	7,024,044	730.0	4526	.47036
Between Full Sisters	589	2,572,827	269.5	4368	.45756
Between 1/2 Sisters	963	4,451,217	460.5	4622	.47819
<hr/>					
$E_P = 4438$		$E_T = .47819$			
$D_P = 547$		$r_{PP1} = .110$		$D_T = .06710$	
		$r_{TT1} = .123$			

^aTitled "Within Dams" in table 18.

The correlation between the temporary environmental variations which affect fat production of maternal sisters would be expected to be small since most of their records would have been made in different years. On the other hand, one would expect some correlation between the permanent environmental effects since maternal sisters are developed in the same uterus even though at different times. Heritability of fat production (intra-herd) would be estimated to be about .320 by assuming $p = 0$; or p would be estimated to be about .043 by assuming that $E_p^2 = .177$, as was estimated from the daughter-dam analysis. Since most of the differences in fat production are environmental, a correlation of only .043 between the total intra-herd

environmental variations affecting each sister would be enough to cause an overestimate of about 81 per cent in the influence of heredity on fat production if g_p^2 actually is .177 but were estimated from the maternal half sister resemblance, ignoring this .043.

The reason for the seemingly close resemblance between maternal half sisters in type ($r_{TT} = .123$) is not clear, considering the correlation of only .074 between dam and daughter and the correlation of .139 between paternal sisters---since the latter are much more contemporary than maternal sisters. From these considerations it seems doubtful that $t = .102$, as is obtained by assuming that $g_T^2 = .142$. On the other hand, an environmental correlation of .102 between maternal sisters does not seem too high if one is willing to accept the possibility that maternal effects (intra-uterine environment, colostrum furnished the calf, etc.) influence the future type of the cow, since many of these effects would be common to maternal sisters.

E. Full Sister Analysis

There were 519 sets of full sisters which were out of 515 different dams and sired by 328 different sires, making a total of 1,108 cows. They were distributed over 180 of the 245 herds which had type ratings. The average number of records per cow was 2.19, and the average production was 453.4 pounds; whereas, the average type was only 4.18.

Since many of the different sets of full sisters within herd were by the same sire, it was necessary to include in the models for type and average production both the sire effect and the effect of the dam within sire, rather than have only one effect which included the combined effects of both parents. Thus, the model for the average production is,

$$\frac{P_{i.klm}}{n_{i.klm}} = \mu + h_i + \sum_j \frac{n_{ijk} y_{ij}}{n_{i.klm}} + s_{ik} + d_{ikl} + e_{iklm}$$

Where:

$\frac{P_{i.klm}}{n_{i.klm}}$ = average fat production of the m^{th} daughter of the l^{th} dam and k^{th} sire in the i^{th} herd.

μ = general mean, characteristic of the population of sets of full sisters.

$\sum_j \frac{n_{ijk} y_{ij}}{n_{i.klm}}$ = year effect remaining in the average production.

s_{ik} = effect of the k^{th} sire in the i^{th} herd.

d_{ikl} = effect of the l^{th} dam mated to the k^{th} sire in the i^{th} herd.

$e_{iklm} = f_{iklm} + \sum_j \frac{e_{ijk}}{n_{i.klm}}$

and

f_{iklm} = effect of the m^{th} daughter of the l^{th} dam and k^{th} sire in the i^{th} herd.

e_{ijklm} = error which is assumed to be random for all records of a given cow and uncorrelated from one cow to another. Assumed $N(0, \sigma_e^2)$.

The model for the type rating of a full sister is,

$$T_{iklm} = \mu + h_l + s_{ik} + d_{ikl} + e_{iklm}$$

Where all subscripts remain the same as in the model for average production.

The s_{ik} effects are assumed to be common for all daughters of the k^{th} sire in the i^{th} herd and uncorrelated from one sire to another. If the sires are related or if the breeders have been successful in selecting sires which have similar genotypes, then the s_{ik} will be correlated and the sire effects will be underestimated.

The d_{ikl} are assumed to be common for all daughters of the l^{th} dam and k^{th} sire in the i^{th} herd and independently distributed from one dam to another within sire and herd. As with the sire effects, independence of the dam effects will not be true if the mates of a sire are more closely related to each other than they are to the mates of other sires in the same herd or if the mates of each sire were intensely selected in a direction not the same as for the mates of other sires in the same herd. Both of these situations are likely to exist in data such as these so that a slight underestimate of the dam effects should

be expected. Still another source of error in estimating the dam effects is the likelihood of a larger correlation between the e_{iklm} between one set of full sisters and another within sire than within sets of full sisters. This also will bias the estimate of the dam component downward.

In the model for the average production the f_{iklm} part of e_{iklm} includes three fourths of the effects of dominance, most of the epistasis, and all of the permanent environmental effects which are characteristic of a cow but not her full sister. These effects can be separated from the temporary environmental effects when records are analyzed individually. They can not be separated from the effects of the temporary environment in the case of type in the present data since only one type rating is available on each cow.

The theoretical compositions of the uncorrected sums of squares and crossproducts and the correction terms are presented in table 21. Note that Y is not separated from E in the crossproducts or sums of squares for type since the type ratings could not be classified by years.

The analyses of variance and covariance and the theoretical composition of the mean squares and covariances for the 519 sets of full sisters are presented in table 22. Y was again assumed to be 515.5, since this analysis affords no estimate of this component. Although it is apparent from the average fat production of full sisters as compared to all cows that they are a select group, there is no reason to suppose that they were

selected in any way that would make the year effects greater or less in their records.

Table 21. Theoretical Composition of Uncorrected Sums of Squares and Crossproducts and the Correction Terms for Type and Average Production of Full Sisters.

Source of Variation	H_P, H_T or $Cov_H(TP)$	Y Average Prod. only	S_P, S_T or $Cov_S(TP)$	D_P, D_T or $Cov_D(TP)$	E_P, E_T or $Cov_E(TP)$
Total	$n \dots$	$\sum_{ijklm} \left(\frac{n_{ijklm}}{n_{i.klm}} \right)^2$	$n \dots$	$n \dots$	$n \dots$
Between Dams	$n \dots$	$\sum_{ikl} \frac{\sum_j \left(\sum_m \frac{n_{ijklm}}{n_{i.klm}} \right)^2}{n_{ikl.}}$	$n \dots$	$n \dots$	Number of Dams
Between Sires	$n \dots$	$\sum_{ik} \frac{\sum_j \left(\sum_{lm} \frac{n_{ijklm}}{n_{i.klm}} \right)^2}{n_{ik..}}$	$n \dots$	$\sum_{ik} \frac{n_{ikl.}^2}{n_{ik..}}$	Number of Sires
Between Herds	$n \dots$	$\sum_i \frac{\sum_j \left(\sum_{klm} \frac{n_{ijklm}}{n_{i.klm}} \right)^2}{n_{i...}}$	$\sum_i \frac{n_{ik..}^2}{n_{i...}}$	$\sum_i \frac{n_{ikl.}^2}{n_{i...}}$	Number of Herds
Correction Term	$\frac{\sum_i n_{i...}^2}{n \dots}$	$\frac{\sum_{ij} \left(\sum_{klm} \frac{n_{ijklm}}{n_{i.klm}} \right)^2}{n \dots}$	$\frac{\sum_{ik} n_{ik..}^2}{n \dots}$	$\frac{\sum_{ikl} n_{ikl.}^2}{n \dots}$	1

The components of variance and covariance as estimated from the analysis on full sisters given in table 22 are presented separately in table 23. The correlations between full sisters

Table 22. Analysis of Variance and Covariance of Type and Average Fat Production of 519 Sets of Full Sisters.

Source of Variation	d.f.	Sum of Sqs. or Crossproduct	Mean Sq. or Cov.	Theoretical Composition of Mean Sq. or Cov.
		<u>P</u>	<u>P</u>	
Intra-Herd Total	928	4,402,075	4,744	$E_P + .339 Y + .781 D_P + .508 S_P$
Between Sires	148	904,696	6,113	$E_P + .541 Y + 2.069 D_P + 3.184 S_P$
Between Dams	191	924,552	4,841	$E_P + .282 Y + 2.190 D_P$
Within Sire and Dam	589	2,572,827	4,368	$E_P + .306 Y$
		<u>T</u>	<u>T</u>	
Intra-Herd Total	928	475.8	.51272	$E_T + .781 D_T + .508 S_T$
Between Sires	148	104.3	.70473	$E_T + 2.069 D_T + 3.184 S_T$
Between Dams	191	102.0	.53403	$E_T + 2.190 D_T$
Within Sire and Dam	589	269.5	.45756	E_T
		<u>TP</u>	<u>TP</u>	
Intra-Herd Total	928	7,088.4	7.6384	$Cov_E(TP) + .781 Cov_D(TP) + .508 Cov_S(TP)$
Between Sires	148	352.2	2.3797	$Cov_E(TP) + 2.069 Cov_D(TP) + 3.184 Cov_S(TP)$
Between Dams	191	1,672.1	8.7545	$Cov_E(TP) + 2.190 Cov_D(TP)$
Within Sire and Dam	589	5,064.1	8.5978	$Cov_E(TP)$

in type and average production and the genetic correlation between those two traits as estimated from these components are given at the bottom of table 23. r_{PP} , and r_{TP} , when adjusted to a single-record basis become .087 and -.037 respectively.

Table 23. Estimates of Variance and Covariance Components from the Full Sister Analysis.

Variance Components		Covariance Components
T	P	TP
$E_T = .45756$	$E_P = 4210$	$Cov_E(TP) = 8.5978$
$D_T = .03492$	$D_P = 222$	$Cov_D(TP) = .07156$
$S_T = .05495$	$S_P = 366$	$Cov_S(TP) = -1.9997$
$r_{TT} = \frac{D_T + S_T}{E_T + D_T + S_T}$	$r_{PP} = \frac{D_P + S_P}{E_P + D_P + S_P}$	
$= .164$	$= .122$	
$r_{TP} = r_{PT} = \frac{Cov_D(TP) + Cov_S(TP)}{\sqrt{E_T E_P}}$		$x = \frac{Cov_D(TP) + Cov_S(TP)}{\sqrt{(D_T + S_T)(D_P + S_P)}}$
$= -.044$		$= -.265$

The environmental contribution to the phenotypic correlations between full sisters (r_{TT} , and r_{PP} ,) was expected to have been about the same as for the maternal half sisters. Only the part of these correlations in maternal half sisters which was genic would be expected to be doubled in the case of full sisters. The fact that r_{TT} , and r_{PP} , for full sisters are much less than

twice what they were for maternal half sisters lends support to the importance of common environmental effects in causing correlations between maternal sisters for both type and production. Since a phenotypic correlation is equal to $g^2 r$ plus $e^2 r_{EE}$, the expected magnitude of r_{TT} and r_{PP} can be calculated for full sisters if g^2 and r_{EE} are known. Assuming that $g_T^2 = .142$, $g_P^2 = .177$, (as observed in the daughter-dam analysis) $t = .102$, and $p = .032$ (as estimated for maternal sisters), then r_{TT} and r_{PP} are expected to be .159 and .105 respectively for full sisters. These agree closely with the correlations observed.

If $\text{Cov}(G_T G_P)$ is actually positive, as evidenced from the daughter-dam and maternal and paternal sister analyses, then the only reasons remaining for $\text{Cov}_S(TP)$ being negative in the full sisters are (1) sampling errors, or (2) one or more of the covariances assumed to be zero in the models were actually negative. Only two of these covariances seem to have a biological interpretation. They are the ones between the effect of the sire for production and the effect of the dam for type and the opposite one. These covariances might conceivably be negative if the breeders were practicing negative assortive mating as they might well be if breeders strive hard to get good type in their sires when they already have outstanding production in their cows but try mainly for high production in their sires when they already have high type in their dams. Even though they were actually making an intense effort to do

this, it is highly unlikely that they were successful enough to account for a negative sire covariance as large as obtained.

In addition, if this were important $\text{Cov}_D(\text{TP})$ should also have been of negative sign.

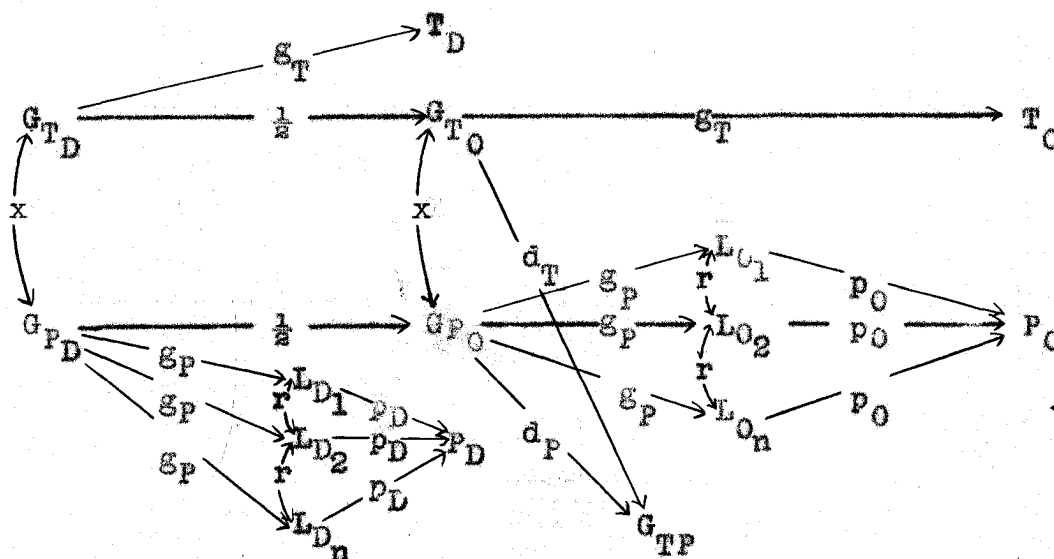
F. Selection for Type and Production

The amount of net genetic improvement possible when selecting for more than one character depends on (1) the intensity of selection for each character, (2) the number of characters being considered, (3) the heritability of each character, and (4) the genetic and phenotypic correlations between the characters. Hazel and Lush (1943) have pointed out that only one combination of selection intensities will allow maximum aggregate gain in a particular set of traits.

Although estimates of the heritabilities and the genetic correlation of type and production were made in this study, no effort was made to determine their relative economic importance. Undoubtedly, the relative economic importance of type and production is not the same in all herds of dairy cattle or even from one year to another within the same herd. It would depend on what the breeder's customers really want. Breeders who dispose of a large portion of their stock through public sales or show ring advertising should pay more attention to type than a commercial dairyman whose income depends almost wholly on his milk or butterfat sales.

By using the genetic and phenotypic constants obtained in the present study and by following the principles of index construction, as given in detail by Hazel (1943), two selection indexes were developed. The first was constructed by giving type one-third as much attention as production and the second, by giving both characters equal weight.

Only information on type and production of the individual being indexed and of her dam were considered in constructing these two indexes. The number of records of fat production were allowed



L_{D_i} and L_{O_i} = the i th lactation record of the dams and daughter respectively.

$$p = \sqrt{\frac{1}{n[1 + (n-1)r]}}$$

$$d_T = a_T \frac{\sigma_{G_T}}{\sigma_{G_{TP}}}$$

$$d_P = a_P \frac{\sigma_{G_P}}{\sigma_{G_{TP}}}$$

and

a_T = the number of pounds of butterfat required to equal in value the difference of one classification grade in type.

a_P = 1 in both indexes.

Figure 4. Path Coefficient Diagram to Illustrate the Relations Between the Aggregate Genotype for Type and Production (G_{TP}) and the Phenotypic Measurements of these two Traits on the Cow being Indexed and her Dam.

to vary from none to four. Figure 4 gives the path coefficient diagram from which was derived the correlations between the phenotypic measurements (T_D , P_D , T_O , and P_O) and the aggregate genotype (G_{TP}) of the daughter which is being indexed. An index (I) is defined

$$I = b_{T_O} T_O + b_{P_O} P_O + b_{T_D} T_D + b_{P_D} P_D$$

where the b 's are multiple regression coefficients which are calculated from the following set of 4 simultaneous equations,

$$\begin{aligned} B_{T_D} + B_{P_D} r_{T_D P_D} + B_{T_O} r_{T_D T_O} + B_{P_O} r_{T_D P_O} &= r_{T_D G_{TP}} \\ B_{T_D} r_{T_D P_D} + B_{P_D} + B_{T_O} r_{P_D T_O} + B_{P_O} r_{P_D P_O} &= r_{P_D G_{TP}} \\ B_{T_D} r_{T_D T_O} + B_{P_D} r_{T_D P_O} + B_{T_O} + B_{P_O} r_{T_O P_O} &= r_{T_O G_{TP}} \\ B_{T_D} r_{T_D P_O} + B_{P_D} r_{P_D P_O} + B_{T_O} r_{T_O P_O} + B_{P_O} &= r_{P_O G_{TP}} \end{aligned}$$

since the B 's are equal to the corresponding b times the ratio of the phenotypic to the aggregate genotype. When the b 's are determined in this manner the multiple correlation between the index and the aggregate genotype ($R_{IG_{TP}}$) is made as large as possible for the economic values considered. Altering the relative economic importance a_T or a_P does not change any of the genetic or phenotypic constants but does alter d_T , d_P , and $\sigma_{G_{TP}}^2$, since

$$\sigma_{G_{TP}}^2 = a_T^2 \sigma_{G_T}^2 + a_P^2 \sigma_{G_P}^2 + 2a_T a_P \sigma_{G_T} \sigma_{G_P} x$$

This in turn changes the correlations between the phenotypic measurements and G_{TP} .

A convenient and practical method of eliminating yearly effects on the production records of an individual cow has not been found. Since selections must be made with these yearly differences present, the genetic and phenotypic constants used in constructing the indexes were those estimated ignoring yearly differences. For convenience the necessary constants are assembled in table 24. σ_P^2 is given for single records of

Table 24. Constants Used to Construct the Two Indexes

Phenotypic Constants	Genetic Constants
$\sigma_P^2 = 7350$	$\sigma_{G_P}^2 = 1229$
$\sigma_P = 85.74$	$\sigma_{G_P} = 35.06$
$\sigma_T^2 = .54299$	$\sigma_{G_T}^2 = .07700$
$\sigma_T = .73688$	$\sigma_{G_T} = .27749$
$r_{TP} = .1174$	$s_P^2 = .1672$
$r_{TD} = .0709$	$s_P = .4089$
$r_{PD} = .0836$	$s_T^2 = .1418$
$r_{PDT} = .0139$	$s_T = .3766$
	$x = .1309$

production. This was calculated by adjusting the estimated intra-herd variance (5,171) from all 8,464 cows to a single-record basis. $\sigma_{G_P}^2$ was then calculated by assuming heritability to be .167 as estimated from the daughter-dam analysis when yearly differences were ignored. The genetic variance for type was calculated in a similar manner. The phenotypic correlations are also given for the case of single records.

Table 25. Partial Regression Coefficients for Type and Average Production of the Dam and Daughter for Different Numbers of Records of Production and the Multiple Correlation Between the Index and the Aggregate Genotype for each Combination, in Two Indexes for Selecting among Daughters.^a

No. Records Dau. Dam	bt _O	bp _O	bt _D	bp _D	Rig _{TP}
<u>Index No. 1</u>					
1 0	6.42	.170	----	----	.412
2 0	5.43	.243	----	----	.479
3 0	4.88	.283	----	----	.513
4 0	4.53	.308	----	----	.534
0 1	----	----	3.21	.085	.206
0 2	----	----	2.72	.125	.240
0 3	----	----	2.44	.141	.257
0 4	----	----	2.27	.154	.267
1 1	6.19	.164	2.70	.071	.446
1 2	6.21	.161	2.28	.102	.458
1 3	6.22	.160	2.05	.119	.464
1 4	6.23	.159	1.90	.130	.468
2 1	5.24	.235	2.73	.065	.505
2 2	5.26	.231	2.35	.094	.514
2 3	5.28	.229	2.13	.109	.519
2 4	5.29	.228	2.00	.119	.522
<u>Index No. 2</u>					
1 0	17.32	.178	----	----	.405
2 0	16.28	.253	----	----	.448
3 0	15.71	.295	----	----	.470
4 0	15.34	.322	----	----	.483
0 1	----	----	8.66	.089	.203
0 2	----	----	8.14	.127	.224
0 3	----	----	7.85	.148	.235
0 4	----	----	7.67	.161	.242
1 1	16.76	.171	7.40	.074	.439
1 2	16.78	.168	6.96	.106	.446
1 3	16.79	.167	6.72	.123	.450
1 4	16.79	.166	6.58	.135	.453
2 1	15.76	.245	7.43	.068	.476
2 2	15.79	.241	7.03	.097	.482
2 3	15.80	.240	6.82	.113	.485
2 4	15.82	.238	6.68	.124	.487

^aIndex No. 1 gives type one-third as much attention as production, whereas Index No. 2 gives type and production equal attention.

The b 's and $R_{IG_{TP}}$'s for the two indexes for selecting a cow when the number of records for her and for her dam vary are given in table 25. In all cases where one or more records of production are available it is assumed here that a type rating is also available. When the amount of information on type and production is the same, $R_{IG_{TP}}$ is always higher in the first index than in the second. This is to be expected since production is estimated to be slightly more heritable than type.

As additional information is added about the producing ability of the dam, the average production of the dam and the type of the cow being indexed are given more weight in the index. The index values are estimates of breeding values after due consideration has been given to the heritability of type and production, the relative economic importance of each trait, and the phenotypic and genetic correlations between the two characters. Thus, the variance of the estimated breeding values,

$$\begin{aligned} \sigma_I^2 = & b_{T_0}^2 \sigma_{T_0}^2 + b_{P_0}^2 \sigma_{P_0}^2 + b_{T_D}^2 \sigma_{T_D}^2 + b_{P_D}^2 \sigma_{P_D}^2 + 2b_{T_0} b_{P_0} \text{Cov}(T_0 P_0) \\ & + 2b_{T_0} b_{T_D} \text{Cov}(T_0 T_D) + 2b_{T_0} b_{P_D} \text{Cov}(T_0 P_D) + 2b_{P_0} b_{T_D} \text{Cov}(P_0 T_D) \\ & + 2b_{P_0} b_{P_D} \text{Cov}(P_0 P_D) + 2b_{T_D} b_{P_D} \text{Cov}(T_D P_D) \end{aligned}$$

will always be less than the variance of the actual breeding values, $\sigma_{G_{TP}}^2$. Progress in actual biological units will therefore be more nearly in proportion to $R_{IG_{TP}}$ than in proportion to $R_{IG_{TP}}^2$.

The amount of gain to be expected, in terms of the aggregate genotype, when using one of the indexes in table 25, is

$$G_{TP} - \bar{G}_{TP} = R_{IG_{TP}} \frac{\sigma_{G_{TP}}}{\sigma_I} (I - \bar{I})$$

I should be almost normally distributed for any particular combination of information, so that if selection is by truncation¹

$$G_{TP} - \bar{G}_{TP} = R_{IG_{TP}} \sigma_{G_{TP}} \frac{z}{b}$$

Where z is the height of the ordinate at the point of truncation and b is the fraction saved for breeding.

Unless the same amount of information is available on all cows being indexed, the type ratings and production records must be expressed as a deviation from some mean for comparison. In small herds, it would probably be best to express the production records as a deviation from a running herd average based on three or four consecutive years, since the year-to-year differences would usually contain more cow differences than in a large herd.¹

An increase in $R_{IG_{TP}}$ could be made by utilizing available information on type and production from still other relatives. However, since most selections of cows are made either before freshening or shortly following their first or second lactations, little is lost by not considering the performance of their

¹Selection by truncation merely means that all individuals above a certain level of merit are retained for breeding and all below that level are discarded.

daughters. Other relatives which might be considered are full sisters and maternal and paternal half sisters. The latter two are likely to be the most abundant but only 25 per cent related to the cow in question; consequently, they will increase the accuracy of the index only a little unless they are numerous and little information is available on the cow herself or her dam.

From a breeder's standpoint it would be desirable to maintain an index for every cow in the herd, utilizing all information that was available. In order to do this accurately the b's must be determined for all possible combinations of number of relatives and records and different kinds of relatives. Relatives which are less than 25 per cent related to the cow being indexed would probably not be worth considering. The b's should also be known for the possible combinations where the type rating is unknown for one or all individuals making up the index. Since the b's vary only a little from one combination to another in most cases, a considerable amount of this difficulty can be overcome by compromising with an index made up with average values for the regression coefficients for type and production. Maximum gain could be more nearly approached by determining the b's for production in only the combinations of kinds and number of relatives with none or one record of production and using as the phenotypic measurement of each individual her most probable producing ability expressed as a deviation from the herd average. The most probable producing

ability has been defined by Lush (1945) to be

$$\text{Herd Average} + \frac{nr}{1 + (n-1)r} \left(\text{Cow's own Average} - \text{Herd Average} \right)$$

where n is the number of records and r is the repeatability of records on the same cow.

V. DISCUSSION

About 8.0 per cent of the differences between single records of production made in the same herd were found in these data to be caused by temporary environmental effects which changed from one year to another. In herds covering 10 years or more the trends and irregular changes in management might have much more important effects on the total variance. This cause of differences between cows is less important, of course, when the average production of all records available for each cow is used. In the present study where the number of records per cow averaged 2.01, yearly differences accounted for only about 4.1 per cent of the total variance between average records of cows kept in the same herd. Yearly differences within herd are thus not a major cause of variation but they are important enough to cause some errors in selections between cows which made their records in different years. Such differences should not be eliminated by expressing each record as a deviation from the herd mean for that year unless the cows which make up the different yearly herd means are mostly the same cows. Otherwise the genetic composition of the herd will change much from one year to another and the actual differences in the yearly means will contain not only environmental differences but many genetic differences as well. Considerable caution should be used in attempting to discount yearly effects by using deviations from the yearly herd means.

Henderson (1949) has suggested a method for estimating yearly effects presumably freed of cow differences so that an individual record can be adjusted to remove effects of yearly changes in general environment. However, the method involves maximum likelihood estimation and a solution of simultaneous equations and thus could not be used readily by a breeder. An accurate and practical method of recognizing and eliminating or discounting the effects of yearly environmental changes is yet to be desired. More investigation is greatly needed along this line.

In the daughter-dam analysis, where the averages of all records available on daughter and dam were used, about 94 and 51 per cent of those yearly effects which were in $\text{Cov}(P_D P_0)$ and σ_{PD}^2 respectively were eliminated by removing the differences between herds. Therefore the regression of .121 for daughter on dam within herd and year was not markedly different from the regression of .119 when years were ignored.

Elimination of yearly effects in order to estimate the correlation between the average production of full sisters, maternal half sisters, or paternal half sisters was not justified in view of the additional labor required and the small difference which resulted in the correlations, as compared to those obtained ignoring yearly effects. Under certain conditions, e.g., if all records of the daughters of a sire were made in the same year and if the data extended over a number of years, it would be quite important to ascertain the influence of yearly

effects on the intraclass correlation between daughters of the same sire. As long as no practical method exists which the breeder can use to eliminate the effects of years, estimates of correlations for use in selection indexes should be made on a basis which neglects year differences.

Utilizing all daughter-dam pairs (4,764) among the 12,405 Jersey cows which had one or more production records Legates (1949) estimated heritability of single records of fat production within herd to be .201. The only restriction on the daughter-dam pairs (2,786) used in this study, where a comparable estimate of .167 was obtained, was that each cow had an official type rating. Since a higher percentage of the herds with better than average production were classified than of those with lower averages, the average production of the daughters and dams used here was about 10 pounds more than for those used by Legates, and, for the same reason, the intra-herd variances were greater in the classified cows. Table 6 seems to indicate that in general breeders were selecting cows which had a type rating as well as one or more production records. Hence, since the genetic correlation is positive, breeders when selecting for type were saving cows which actually had slightly better genotypes and were culling cows which had slightly worse genotypes for production than their phenotypes indicated. As pointed out by Lush (1948), this type of selection tends to decrease the regression of daughter on dam since the dam's breeding value is better than her phenotype indicates.

This source of bias would reduce the regression for type as well as the regression for production. The importance of this bias is probably slight since the genetic correlation between type and production seems to be only about .15 to .20.

The intra-herd regression of type of daughter on type of dam was .071 in this study. The effect of differences between judges in herds which were classified more than once could not be determined since the judges were not listed. Tyler and Hyatt (1948) found an intra-sire regression of .14 in Ayrshires where the daughter and dam were classified by the same judge on the same day. The difference between their estimate and the present one may indicate that differences between judges are more important in these Jersey data than one would suppose from a priori grounds, or that the heritability of differences in official type ratings is considerably lower in Jerseys than in Ayrshires. The estimate of heritability of intra-herd differences in type ratings for actual use in selection indexes should be made with the judge differences present, unless a satisfactory method is found for breeders to eliminate them. Differences between judges should not, as a rule, be eliminated by expressing the type rating as a deviation from the mean for that judge for the same reason that yearly effects in production records should not be eliminated by expressing them as a deviation from the yearly herd mean, i.e., genetic differences are also eliminated. If the yearly or judge-to-judge differences in type ratings are found to be an extremely large element in the total variance of

those, it might be a good compromise to sacrifice some of the genetic differences in order to get rid of the judge differences.

The influence of judge and year differences on the regression of type of daughter on type of dam needs much further investigation. In the two studies that have been made on variations of type ratings of the same cow, Johnson and Lush (1943) with Holsteins classified at 12 month intervals and Hyatt and Tyler (1948) with Ayrshires classified at 4 to 6 month intervals, the correlation between type ratings made on the same cow by different judges was estimated to be .34 and .55, respectively. The repeatability in the latter study was higher (.62 to .82) when the same judge rated the cow at different times. How much of this increase was due to the judge remembering his former ratings or to real differences in ideals of judges could not be determined. The interval between ratings was much shorter in the Ayrshire data. No comparable study has been reported in Jerseys.

The intra-sire and intra-herd variance of type ratings was .478 as compared to an intra-sire and intra-inspector variance in Ayrshires of .667, as reported by Tyler and Hyatt (1948), when a comparable coding system of type ratings was used. This seems to indicate that either (1) paternal half sisters in Jerseys are more uniform in type, or (2) that the official judges which classified the Jerseys were more conservative and thus failed to utilize the extreme classes as much as those which classified the Ayrshires. The average type of the 8,464 cows used in this

study was 4.19 as compared to an average of 4.13 for 5,177 Ayrshire cows studied by Tyler and Hyatt which had production records.

The phenotypic covariance between type and production $[Cov_E(TP)]$ in a population of unrelated cows contains one component of genetic covariance $[Cov(G_T G_P)]$ and one component of environmental covariance $[Cov(E_T E_P)]$. In a population of half sisters it still contains all of $Cov(E_T E_P)$, but only $3/4$ of $Cov(G_T G_P)$; whereas, in a population of full sisters $1/2$ of $Cov(G_T G_P)$ disappears. Within sets of identical twins $Cov(G_T G_P)$ would disappear entirely and the only reason for type and production to be correlated on the same individual is common environmental effects, since the genetic crossproduct is the same for each twin. The correlations between the environments for type and production of identical twins would be fairly high so that $Cov(E_T E_P)$, $\sigma_{E_T}^2$, and $\sigma_{E_P}^2$ would be smaller than in an unrelated population. $\sigma_{E_T}^2$ and $\sigma_{E_P}^2$ should reduce proportionately more than $Cov(E_T E_P)$, thus, the phenotypic correlation between type and production would probably be about the same as in an unrelated population.

Breeders of dairy cattle have often maintained that selection for type would also bring about improvement in production. In general there seems to be some truth in this. However, if selection is based solely on the type of the cow the genetic improvement in production will be only about $1/6$ as fast ($r_{TG_P} = .068$ and $r_{PG_P} = .409$) as selection based solely on one

production record of the cow in question. If type has no economic value at all, the genetic gain in production would be only about one-half of one per cent faster by considering type as well as production. A type rating on the dam and one on the cow being considered for selection is about 18 per cent as valuable in predicting her breeding value for production as one of her own production records, or 36 per cent as valuable as one production record on the dam when no records of production are available on the daughter or dam.

VI. SUMMARY

The most reliable estimates of the contribution of the different sources of variance and covariance in official type ratings (one rating per cow) and average fat production (average of two records per cow) were as follows:

Variance or Covariance Caused by	Type	Average Production	Covariance
Herds	.1167	3939	4.110
Cows within Herds	.5430	5171	7.419
Genic	.0770	1229	1.769
Other	.4660	3942	5.650

These variance and covariance components were calculated from tables 9, 10, and 15. The type ratings were coded so that a difference of one unit separated two adjacent classes.

Genetic variation constituted 18 per cent of the intra-herd and intra-year variance of single records of fat production in a population of Jersey cows who all had an official rating for type. Intra-herd differences in type ratings were 14 per cent heritable. The genetic correlation between type and production was estimated as .18 from 2,786 daughter-dam pairs, .12 from the paternal sister analysis, .21 from the maternal sister analysis, and $-.26$ from the 519 sets of full sisters. The paternal and maternal sister analyses contained the same 8,464 cows.

The intra-herd phenotypic correlation between type and a single production record of the same cow was .12. Only about 15 to 30 per cent of this correlation was caused by genes which

affected both type and production alike; the remainder was caused by environmental variations which affected both alike.

On the basis of single records, the intraclass correlations between the phenotypes of half and full sisters were as follows:

Correlation	Half Sisters		Full Sisters
	Paternal	Maternal	
$r_{TT'}$.14	.12	.16
$r_{PP'}$.12	.08	.09
$r_{TP'} = r_{PT'}$.02	.03	-.04

In view of the estimates of heritability for type and production from the daughter-dam analysis, the environmental contribution to $r_{TT'}$ and $r_{PP'}$ for paternal half sisters was more important than the genetic. If the environmental correlations for paternal half sisters are largely a result of contemporaneity, then the environmental correlations for maternal half sisters should be considerably smaller unless the intra-uterine environmental effects are important and correlated. Since $r_{TT'}$ was almost as large for maternal half sisters as for paternal half sisters and $r_{PP'}$ was considerably smaller, there is some indication that the intra-uterine environment may influence the future type of the cow more than it does her future production. $r_{TP'}$ and $r_{PP'}$ for full sisters were about as expected considering the heritability estimates from the daughter-dam analysis and the expected environmental contributions. $r_{TP'}$ was less than expected for full sisters but the degrees of freedom are few enough that sampling errors could account for this discrepancy.

Differences in average type from one herd to another accounted for 16 per cent of the total variance in type ratings. Differences between herds were considerably more important in the total variance of average production, accounting for 40 to 45 per cent of that. No effort was made to ascertain the importance of heredity and environment in causing the differences between herds, either in type or in production.

The fraction of the intra-herd variance in average production due to yearly changes in the general environment of that herd was only 4.1 per cent. The yearly effects on production are largely eliminated when the average of all records on the cow is studied. The amount remaining in the average of n records will be about $1/n$ of what it was for single records. The distribution of these yearly effects with respect to the mean squares used to determine components of variance for calculating correlations between the average production of paternal and maternal sisters or full sisters in dairy data is such that they can usually be ignored; unless one is interested in the magnitude of the various components rather than the ratio which is necessary to calculate the correlations.

A comparison of the average type rating of 2,044 dams, which had a daughter with one or more records of production, with the average type of all 8,464 cows indicated that breeders were giving some attention to type in their selection program. Proportionately, the difference between the averages of dams and of daughters was larger for type than production. Two possible

explanations are given for this difference : (1) the dams were older than the daughters and age differences were not removed from the type ratings, and (2) breeders actually may have been giving more attention to type than production in their selection of the dams.

Using the heritability estimates, genetic correlation, and phenotypic correlations obtained in this study, two selection indexes were developed by Hazel's multiple regression technique; first, by giving type one-third as much attention as fat production and, second, by giving both characters equal attention. Only information about the phenotypes of the dam and her daughter were considered in constructing these indexes. Although selection on the basis of type alone should automatically bring about some genetic improvement in production, it would require about 6 generations to obtain the improvement that selection on the basis of production would obtain in one generation. The most efficient gain is secured by using a selection index that has been constructed properly and utilizes all information that might be available on both type and production. Regression coefficients for some frequently met combinations of information about a cow and her dam are shown in table 25. Several of these combinations will yield progress about half as fast ($R_{IG_{TP}}$ is about .5) as if the exact Mendelian genotype of the cow were known.

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.
- Copeland, Lynn. 1938. The old story of type and production. Jour. Dairy Sci. 21:295-303.
- Copeland, Lynn. 1941. The relationship between type and production. Jour. Dairy Sci. 24:297-304.
- Gowen, J. W. 1920. Conformation and its relation to milk producing capacity in Jersey cattle. Jour. Dairy Sci. 3:1-32.
- Gowen, J. W. 1921. Studies on conformation in relation to milk producing capacity in cattle. II. The personal equation of the cattle judge. Jour. Dairy Sci. 4:359-374.
- Gowen, J. W. 1924. Milk secretion. Williams & Wilkins Co., Baltimore, Maryland.
- Gowen, J. W. 1934. The influence of inheritance and environment on the milk production and butterfat percentage of Jersey cattle. Jour. Agr. Res. 49:433-465.
- Hazel, L. N. 1941. Principles of a selection index which involves several characteristics and utilizes information concerning relatives. Unpublished Ph. D. thesis. Library, Iowa State College, Ames, Iowa.
- Hazel, L. N. 1943. The genetic basis for constructing selection indexes. Genetics 28:476-490.
- Hazel, L. N., Baker, M. L., and Reinmiller, C. F. 1943. Genetic and environmental correlations between the growth rates of pigs at different ages. Jour. Anim. Sci. 2:118-128.
- Hazel, L. N. and Lush, Jay L. 1943. The efficiency of three methods of selection. Jour. Hered. 33:393-399.
- Henderson, C. R. 1949. Estimation of changes in herd environment. Paper presented at the forty-fourth annual meeting of the American Dairy Science Association. (Abs.) Jour. Dairy Sci. 32:706.
- Hyatt, George, Jr. and Tyler, W. J. 1948. Variations in type ratings of individual Ayrshire cows. Jour. Dairy Sci. 31:71-79.

- Hyatt, George, Jr., Tyler, W. J., and Conklin, C. T. 1949. The relationship between type ratings of Ayrshire females as young heifers and as cows. Jour. Dairy Sci. 32:375-380.
- Johansson, Ivar and Hansson, Artur. 1940. Causes of variation in milk and butterfat yield of dairy cows. Kungl. Lantbruksakademiens Tidskrift 79, Specialhäfte 3:1-127.
- Johnson, Leslie E. and Lush, Jay L. 1942. Repeatability of type ratings in dairy cattle. Jour. Dairy Sci. 25:45-56.
- Legates, J. E. 1949. A selection index for butterfat production in Jersey cattle utilizing the fat yields of the cow and her relatives. Unpublished Ph. D. thesis. Library, Iowa State College, Ames, Iowa.
- Lush, Jay L. 1940. Intra-sire correlations or regressions of offspring on dam as a method of estimating heritability of characteristics. Proc. Amer. Soc. Anim. Prod. 1940: 293-301.
- Lush, Jay L. 1945. Animal breeding plans. 3rd ed. The Collegiate Press, Inc., Ames, Iowa.
- Lush, Jay L. 1948. The genetics of populations. Unpublished mimeographed notes.
- Lush, Jay L., Norton, H. W., III., and Arnold, Floyd. 1941. Effects which selection of dams may have on sire indexes. Jour. Dairy Sci. 24:695-721.
- Lush, Jay L. and Straus, F. S. 1942. The heritability of butterfat production in dairy cattle. Jour. Dairy Sci. 25:975-982.
- Plum, Mogens. 1934. Production in a large Jersey herd as affected by sires, dams and yearly variations. Proc. Amer. Soc. Anim. Prod. 1933:53-57.
- Plum, Mogens. 1935a. Causes of differences in butterfat production of cows in Iowa Cow Testing Associations. Jour. Dairy Sci. 18:811-825.
- Plum, Mogens. 1935b. The relative importance of heredity and environment in determining the butterfat production of cows in Iowa Cow Testing Associations. Unpublished Ph. D. thesis. Library, Iowa State College, Ames, Iowa.
- Satterthwaite, F. E. 1946. An approximate distribution of estimates of variance components. Biometrics Bull. 2:110-114.

- Snedecor, G. W. 1946. Statistical methods applied to experiments in agriculture and biology. 4th ed. The Collegiate Press, Inc., Ames, Iowa.
- Touchberry, Robert W. 1948. Genetic correlations in five body measurements, weight, type, and production of Holstein cows. Unpublished Ph. D. thesis. Library, Iowa State College, Ames, Iowa.
- Tyler, W. J. and Hyatt, George, Jr. 1947. The heritability of milk and butterfat production and percentage of butterfat in Ayrshire cattle. (Abs.) Jour. Anim. Sci. 6:479-480.
- Tyler, W. J. and Hyatt, George, Jr. 1948. The heritability of official type ratings and the correlation between type ratings and butterfat production of Ayrshire cows. Jour. Dairy Sci. 31:63-70.
- Wright, Sewall. 1921. Correlation and causation. Jour. Agr. Res. 20:557-585.
- Yates, F. 1933. The principles of orthogonality and confounding in replicated experiments. Jour. Agr. Sci. 23:108-145.

VIII. ACKNOWLEDGMENTS

Words are inadequate to express the heartfelt thanks and appreciation for the many helpful suggestions and constructive criticisms received from Dr. J. L. Lush during the entire course of the author's graduate study, and particularly with problems involved in preparing this manuscript.

The author is also indebted to Dr. L. N. Hazel for his assistance in many of the statistical problems of this investigation and for his constructive criticisms of the manuscript.

Last but not least goes thanks and appreciation to the American Jersey Cattle Club for the assistance which helped make this investigation possible; and to its Research Director, Mr. John P. Beardsley, for his efforts in providing the necessary data in the form desired.

IX. APPENDICES

Appendix A.

Summary of the Number of Cows with
Type and Production Records
by Herds and States

Alabama

74 C. C. Glen & Sons, Decatur

Arkansas

29 Ark. Agric. Expt. Sta., Fayetteville

California

27 Estate of J. W. Coppini, Ferndale
27 N. C. Bowles, Eureka
24 M. Barcelles & Son, Arcata
166 E. E. Greenough, Merced
51 Guy H. Miller, Modesto
24 A. J. Regli, Ferndale
25 California Poly. School, San Louis Obispo
44 Vernon Thornburg, Turlock
28 John Giacomini, Eureka
22 Mr. and Mrs. D. F. Anderson, Ft. Lupton
40 Mr. and Mrs. W. J. Mack, Pleasanton
478

Connecticut

36 Ellen M. Colman, Pomfret
21 C. F. Buell, Woodstock Valley
61 Mrs. Waldo S. Kellog, Derby
23 Univ. of Conn., Storrs
141

Florida

88 Walter Welkener, Jacksonville

Georgia

53 H. D. Allen, Milledgeville
81 Univ. of Georgia, Athens
22 Georgia Experiment Station
45 Ga. Coastal Plain Expt. Sta. Tipton
24 J. J. Harris, Pelham
228

Idaho

- 21 H. W. Mc Cavley, Buhl
- 22 F. M. Bledsol, Caldwell
- 41 Univ. of Idaho, Moscow
- 84

Illinois

- 16 Allen Britton, Marseilles
- 21 Floyd H. Phillips, Barry
- 25 Vernon O'Brien, Mendon
- 21 V. T. Winnings, Lorington
- 11 R. W. Riggs, Jacksonville
- 25 H. P. Bicknell, Lovington
- 7 J. C. Franklin, Leroy
- 23 H. A. Gardner, Wadsworth
- 20 S. R. Griffin, Clinton
- 40 Chester J. McCord, Newton
- 13 J. H. McCutcheon, Springfield
- 29 Earl Tenhouse, Liberty
- 37 O. L. Putnam, Harvard
- 288

Indiana

- 38 L. O. Miller, Goshen
- 16 P. W. Canary, Franklin
- 24 R. D. Gonser, Hudson
- 18 P. W. Mc Kibben & Son, Orland
- 35 Walter E. Wolfe, Indianapolis
- 131

Iowa

- 13 J. R. Jose & Son, Stuart
- 3 P. V. Durnan, Ossian
- 17 J. O. Bartz, Rudd
- 28 Barr Keshlear, Shenandoah
- 45 W. S. Morrison, Iowa City
- 36 J. N. Gates, West Liberty
- 22 Hi Bred Corn Co., Johnston
- 5 Elmer Nabholz & Sons, Indiana
- 17 Iowa State College, Ames
- 19 Claire Wemer, Keswich
- 21 Graceland College Farms, Lamoni
- 26 Bert W. Sams & Sons, Beaman
- 252

Kentucky

- 31 J. Camenisch & Sons, Stanford

- 42 C. W. Shipman, Shelbyville
- 81 P. B. Gaines, Carrollton
- 41 Kentucky Agric. Expt. Sta., Lexington
- 33 F. Rouse & Sons, Crittendon
- 228

Louisiana

- 97 La. State Univ., Baton Rouge

Maine

- 53 Pine Hill Farms, Waterville
- 13 George Meylan, Cosco
- 66

Maryland

- 49 J. W. and H. G. Henry, Northeast
- 48 William Fahnestock, Jr., Centreville
- 21 W. P. Hoopes, Forest Hill
- 53 Herbert Hoopes & Son, Bel Aire
- 171

Massachusetts

- 24 Mrs. Sidney Howard, Tyringham
- 15 R. L. Manchester, Amherst
- 12 Mass. State College, Amherst
- 173 Mr. and Mrs. H. G. Wilde, Lenox
- 38 K. M. Stevens, North Arlington
- 87 J. R. Sibley, Spencer
- 56 A. F. Means, Brookfield
- 405

Michigan

- 15 Cecil J. Baker & Sons, Lenon
- 16 A. D. Mac Rae, Mt. Pleasant
- 19 Emmett H. Davis, Rockford
- 21 Michigan State College, East Lansing
- 25 J. A. Cronkright, Weberville
- 8 D. S. and W. O. Sneed, Durand
- 11 Oscar Kaechele, Middleville
- 58 North Ridge Jersey Farms, Chesaning
- 9 W. L. Ayers, Jasper
- 57 C. M. Brown, Tecumseh
- 36 J. K. Hatfield & Sons, Remus
- 10 Ralph H. Schroeder, Owasso
- 6 Warren H. Clay, Wayland
- 18 Arden W. Wright, Marcellus

- 15 K. L. Rorison, Goodrich
- 26 Hayes Jersey Farm, Clarklake
- 366

Minnesota

- 43 C. L. Ames, St. Paul
- 31 Univ. of Minn., St. Paul
- 51 F. B. Astroth, St. Paul
- 25 A & D Lippert, Bertha
- 27 Crescent Ridge Farm, Minneapolis
- 11 St. Mary's Mission School, Redlake
- 38 A. A. Haberdank, Ada
- 226

Missouri

- 57 J. K. Burney, Aurora
- 54 P. R. Smith, Joplin
- 14 Ralph C. & P. W. McDaniel, Versailles
- 60 W. N. Burney, Farm No. 1, Aurora
- 21 J. R. Hearne, Noel
- 7 W. L. McAllister, Shelbyna
- 32 Hatch Dairy Expt. Sta., Oakwood
- 35 W. N. Burney, Farm No. 2, Aurora
- 19 Shearman Miller, Marshall
- 20 Univ. of Missouri, Columbia
- 29 J. E. Aston, Rolla
- 56 The School of the Ozarks, Point Lookout
- 36 W. J. Knotwell, Norwood
- 33 Jewell Windle, Springfield
- 28 M. Saffarraus, Palmyra
- 14 Lee Heselton, West Plains
- 515

Nebraska

- 11 Univ. of Nebraska, Lincoln

New Hampshire

- 51 Robert P. Bass, Peterborough
- 33 Univ. of New Hampshire, Durham
- 84

New Jersey

- 4 N. Kelly, Annandale
- 44 Lila Tyng, Gladstone
- 14 Herman J. Lazarus, Pillstown
- 40 Forwood Hills Farms, High Ridge
- 69 Marlu Farm, Lincroft

- 76 Caleb S. Ridgeway Jr., Columbus
- 6 N. J. Agric. Expt. Sta., New Brunswick
- 64 P. D. Van Mater, Malboro
- 53 J. F. O'Brien, Lebanon
- 120 R. O. Gubelman, Whitehouse
- 41 Duke Farms, Somerville
- 531

New Mexico

- 11 Eastern New Mexico College, Fortales
- 20 New Mexico Coll. of Agr. and Mech., State College
- 31

New York

- 25 Lee Chamberlain, Albion
- 48 Ira G. Payne, East Schadack
- 98 W. R. Kenan Jr., Lockport
- 10 Cornell University, Ithaca
- 89 S. M. Chettenden, New Lebanon
- 270

North Carolina

- 35 J. O. Lutz, Newton
- 38 N. C. State College of Agric., Raleigh
- 34 E. L. Moose, Conover
- 107

Ohio

- 64 Borden's of Springfield, Springfield
- 39 Finch Jersey Farm, Dayton
- 18 Broughton's Farm, Marietta
- 22 F. E. Rathburn, Renoldsburg
- 24 Russell Hoar, Newark
- 20 J. J. Barth, Cleveland
- 31 Billman Bros., Burbank
- 7 Mrs. Harriet R. Keller, Glendale
- 31 Chester Folck & Sons, Springfield
- 10 Floyd Damschroder, Gibsonburg
- 45 A. E. Stevens, Loveland
- 89 F. H. Grant, Centerville
- 28 L. G. Haverland, Hamilton
- 17 W. D. Kahler, Westerville
- 22 Glenn M. Lackey & Sons, Delaware
- 42 Fred S. Schrock, Worthington
- 22 Ohio State Univ., Columbus
- 5 Wm. J. Young & Son, Troy
- 7 C. R. Kessler, Greenville
- 32 Alexander & Trout, Spring Valley

27 High Hollow Farms, Lebanon
8 C. S. Gibson, Quaker City
610

Oklahoma

65 C. E. Hiatt, Braman

Oregon

13 C. J. & L. J. Hunter, Ashland
34 Ralph & Arthur Knox, Gaston
52 George J. Hornings Jr., Sherwood
36 E. B. Payer & Sons, Ashland
20 Oregon State Agric. Coll., Corvallis
47 R. E. Cope, Langlois
50 Mikkelsen & Son, Woodburn
18 D. R. Dickie, Newberg
44 L. S. Lorenzen, Dayton
38 S. U. & M. K. Drages, McMinnville
30 J. F. Buiperie & Son, Woodburn
34 S. W. Melott, Portland
18 J. E. Lindow, Independence
37 Frank Schutzwall, Grant's Pass
11 Mrs. Elmer Miller, Siletz
20 R. W. Williams & Family, Carlton
20 Chester J. Jenkins, Independence
42 E. E. Gourley, Albany
27 G. A. Gentemann, Independence
23 Robert Atkinson, Sand Lake
614

Pennsylvania

45 R. N. Moose & Sons, Mercer
32 Hill Farm, Coatesville
26 Raymond Antonin, New Hope
49 C. E. Massinger, Phoenixville
36 William A. E. Leitzinger, Clearfield
77 Mr. and Mrs. J. S. Campbell Jr., Butler
22 Penn. State College,
237

Rhode Island

57 Edith Howe De Wolf, Bristol
12 R. I. State College, Kingston
69

South Carolina

35 Ward Crim, Moore
16 E. L. Mc Kown, Gaffney
51

Tennessee

- 30 Sample Bros., Rutherford
- 34 Robert Jordan, Franklin
- 123 Shelby Co. Penal Farm, Memphis
- 19 John Jr. & J. E. Thomason, Nolenville
- 206

Texas

- 27 Earl Barber, Alvin
- 11 Texas Tech. College, Lubback
- 32 Victory Jersey Farm, Tulia
- 10 Orrland Jersey Farms, Tulia
- 59 Emil Reus, San Antonio
- 22 R. F. Mc Patrige Jr., Big Sandy
- 82 The Luling Foundation, Luling
- 46 Texas Agric. and Mech. College, College Station
- 81 Jim Maverick, San Antonio
- 11 J. C. White, Rotan
- 381

Utah

- 13 W. J. Rasband, Helier
- 23 Utah State Agric. College, Logan
- 8 J. Sylvan Rasband, Helier
- 44

Vermont

- 33 G. O. Neil, Woodstock
- 45 W. D. Perry, Hartland Four Corners
- 169 E. S. Brigham, St. Albans
- 19 State School of Agriculture, Randolph Center
- 30 W. C. & F. Arms, Burlington
- 34 Lexington Farm, Reading
- 26 J. & C. L. Bellefeville, Tunbridge
- 12 Vt. Agric. Expt. Sta., Burlington
- 5 D. H. Wright & Sons, West Brattleboro
- 373

Virginia

- 14 Clover Hill Farm, Manassas
- 48 W. A. Williams, Orange
- 21 J. S. Roller, Timberville
- 40 J. S. Andrews, Orange
- 123

Washington

- 10 Leroy F. Caverly, Monroe

14 Cliff Henning, East Stanwood
39 E. D. McPherson, Enumclau
104 W. F. Nelson, Walla Walla
6 J. J. Eason, Bellingham
41 T. L. Fishback, Chehalis
19 Walter Gustafson, Rochester
12 Rodney E. & Charles E. Olson, Mt. Vernon
71 Chas. H. Winell Inc., Shelton
316

West Virginia

27 C. A. Woodworth, Hurricane
70 J. A. Kelley, Huntington
97

Wisconsin

15 Univ. of Wisconsin, Madison
32 C. B. Finn, Bridgeport
52 Steffanus, Delavan
38 Chas. S. Kelley, Hudson
26 Lee Wilkins, Livingston
163

Appendix B.

An Algebraic Comparison of the Number of Sire Components
in the Various Sums of Squares and Mean Squares with
Proportionate and Disproportionate Frequencies

The linear mathematical model is

$$Y_{ijk} = \mu + h_i + s_{ij} + e_{ijk}$$

Where:

Y_{ijk} = the k^{th} observation of the j^{th} sire in the i^{th} herd.

μ = general mean

h_i = effect of the i^{th} herd. $i = 1, 2, \dots, m$

s_{ij} = effect of the j^{th} sire in the i^{th} herd.
 $j = 1, 2, \dots, q$

e_{ijk} = error associated with k^{th} observation.
 $k = 1, 2, \dots, p$

mqp = total number of observations = $n \dots$

Proportionate Frequencies

Disproportionate Frequencies

Correction Term

$$\frac{mqp^2}{mqp} = p$$

$$\frac{\sum \frac{n_{.j}^2}{n_{...}}}{n_{...}}$$

Between Sires

Uncorrected
Sum of Squares mqp

$n_{...}$

Corrected $mqp - p$
Sum of Squares $= p(mq - 1)$

$$n_{...} - \frac{\sum \frac{n_{.j}^2}{n_{...}}}{n_{...}}$$

Mean Square $\frac{p(mq - 1)}{mq - 1}$

$$n_{...} - \frac{\sum \frac{n_{.j}^2}{n_{...}}}{n_{...}}$$

$= p$

$$\frac{\sum \frac{n_{.j}^2}{n_{...}}}{\sum \frac{n_{.j}^2}{n_{...}} - 1}$$

Between Herds

$$\begin{aligned} \text{Uncorrected} & \quad \frac{mqp^2}{qp} \\ \text{Sum of Squares} & \\ & = mp \end{aligned}$$

$$\sum_i \frac{\sum_j n_{ij}^2}{n_{i..}}$$

$$\begin{aligned} \text{Corrected} & \quad mp - p \\ \text{Sum of Squares} & = p(m - 1) \end{aligned}$$

$$\sum_i \frac{\sum_j n_{ij}^2}{n_{i..}} - \frac{\sum_j n_{.j}^2}{n_{...}}$$

$$\begin{aligned} \text{Mean Square} & \quad \frac{p(m - 1)}{m - 1} \\ & = p \end{aligned}$$

$$\frac{\sum_i \frac{\sum_j n_{ij}^2}{n_{i..}} - \frac{\sum_j n_{.j}^2}{n_{...}}}{m - 1}$$

Between Sires Within Herd

$$\begin{aligned} \text{Corrected} & \quad mqp - mp \\ \text{Sum of Squares} & = mp(q - 1) \end{aligned}$$

$$n_{...} - \sum_i \frac{\sum_j n_{ij}^2}{n_{i..}}$$

$$\begin{aligned} \text{Mean Square} & \quad \frac{mp(q - 1)}{mq - 1} \\ & = p \end{aligned}$$

$$\frac{n_{...} - \sum_i \frac{\sum_j n_{ij}^2}{n_{i..}}}{\sum_{ij} n_{ij} - m}$$

Within Herd

$$\begin{aligned} \text{Corrected} & \quad mqp - mp \\ \text{Sum of Squares} & = mp(q - 1) \end{aligned}$$

$$n_{...} - \sum_i \frac{\sum_j n_{ij}^2}{n_{i..}}$$

$$\begin{aligned} \text{Mean Square} & \quad \frac{mp(q - 1)}{m(qp - 1)} \\ & = \frac{p(q - 1)}{qp - 1} \end{aligned}$$

$$\frac{n_{...} - \sum_i \frac{\sum_j n_{ij}^2}{n_{i..}}}{n_{...} - m}$$