

VARIATION IN BUTTERFAT PRODUCTION  
IN STATE-OWNED HERDS IN IOWA

by

Devadason Sundaresan

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

Major Subject: Animal Breeding

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University  
Of Science and Technology  
Ames, Iowa

1959

## TABLE OF CONTENTS

	Page
INTRODUCTION	1
HISTORICAL	5
DATA	10
Description	10
Tests	14
METHOD OF INVESTIGATION	22
RESULTS	29
Analysis of Variance Technique	29
Findings on Environmental Factors	33
Components	33
Herd effect	42
Year effect	45
Season effect	48
Interactions	51
Findings on Genetic Factors	58
Hierarchal classification	60
Full sister analysis	66
Cross classification	70
Confidence limits of components	72
DISCUSSION OF RESULTS	75
SUMMARY	90
LITERATURE CITED	93
ACKNOWLEDGMENTS	99

## INTRODUCTION

From ancient times man has domesticated and raised cattle to provide food as milk and meat, or to be a beast of burden. Milk production is a physiological process necessary for the survival of the species. Natural selection must have operated on this characteristic to ensure production of an amount of milk adequate for the young. When man wanted to increase production of milk, so that he could use for himself the milk in excess of the needs of the calf, he initiated artificial selection and planned mating. Presumably he also attempted to offer an environment that is conducive to increased milk production. One gets the impression, from references made in early literature on livestock breeding, that his breeding program, wherever it deviated from random mating, had a strong reliance on the old adage, "like begets like."

Milk production, as exhibited by an individual cow, is the response of its genotype for this characteristic under the particular environment in which the individual is raised and makes the record. With characteristics like milk production which are highly influenced by environmental differences, the average cattle breeder cannot be as accurate in his selections as he can be with characters less affected by environment. The average genotype of a population changes only slowly within short periods of time.

During the second half of the nineteenth century organized effort had been directed to improving the practice of agriculture in Europe and U.S.A. In the latter country, through the establishment of Land Grant Colleges, considerable attention was directed to livestock feeding and management. During these early years, while some attention had been given to the practice of animal breeding, the major experimental emphasis had been on problems of feeding and management. This is understandable since breeding studies are expensive and results are obtained after a considerable period of time. Early livestock genetics was mostly restricted to the analysis of simply inherited characters in an individual, and to identifying genes with major phenotypic effects. Mendelian genetics later was extended to characters in a population affected by a number of independent factors. During the last few years, milk production per cow has increased both in Europe and in North America. According to U.S. Agricultural Marketing Service (1956), the average milk production per cow in U.S.A. in 1924 was 4,167 pounds in contrast to 5,657 pounds in 1954. This advancement has surely been brought about by improving both the genetic merit of cattle and their feeding and management. One cannot specify certainly the relative role of these factors.

Earlier breeding techniques have been essentially mass

selection with an uncertain amount of attention to pedigrees, performance of relatives, and selection and mating within the same "pure breed." If heritability is high for a certain characteristic, mass selection is quite effective. Lush (1940) defined heritability as the fraction of the observed variance which was caused by differences in heredity. This statistic is obtained indirectly from differences observed in the particular characteristic in related animals. The statistic thus obtained, in most cases, is a measure of average effects of the genes involved. For characteristics with low heritability mass selection is relatively ineffective. It is desirable then to resort to other breeding techniques, such as using family averages in addition to the individual's own characteristic in predicting the breeding value of that individual. With increased need of complex breeding techniques, it is desirable to get more reliable estimates of the different sources of variation in milk production under different circumstances. It is also desirable to estimate the relative importance of non-additive gene effects. The purpose of this investigation is to estimate different sources of variation in milk production, using some fairly extensive data available from the state-owned herds in Iowa, a population which may be somewhat

different from D.H.I.A. farm herds. The earlier work in U.S.A. on this subject is mostly from D.H.I.A. or H.I.R. or Agricultural Experiment Station herds.

## HISTORICAL

Darwin (1875, pp. 82-97) gives evidence of domestication of cattle as early as 2100 B.C. In this work he dealt with evolution of breeds, variation among them, and how breeds had been modified by methodical selection. He believed that characters thus modified were strongly inherited. This gives room for the suggestion that Darwin was aware that performance of a character was not solely determined by heredity. The fusion or blending theory of inheritance was the accepted belief then, which was shared by Darwin. However, Fisher (1958, pp. 1-2) gives evidence that Darwin saw the need for an alternative concept to blending inheritance though he never worked out any specific theory.

Irrespective of the stage of theoretical knowledge in inheritance, Darwin's work (1875, p. 447) gives evidence that the art of breeding was highly successful with several species of animals. He refers to pedigrees of game cocks, greyhounds, pigs and cattle, and the high prices paid for prize animals. The use of pedigree in breeding can be traced back to the eighteenth century, while there is historic evidence that this was practiced even earlier. Breed registry societies for cattle, a necessary institution for widespread pedigree breeding, came into existence early in the nineteenth century. An adequate review of early animal breeding is found in the first five chapters of "Animal

Breeding Plans" by Lush (1945). Mendelism was rediscovered at the turn of the century. The first few years the geneticists were mainly concerned with verifying Mendelism with various livestock. In 1919, East and Jones (1919, p. 50) contended that "Mendelian heredity has proved to be the heredity of sexual reproduction; the heredity of sexual reproduction is Mendelian!" According to Lush (1958) the Mendelian discoveries did little to change the art of breeding followed by breeders of livestock but offered an explanation for the puzzling situations, such as reversion, regression toward breed average, consequences of inbreeding, etc.

The application of Mendelian genetics to populations, in biometrical terms considered the beginning of "population genetics" could be attributed to Hardy (1908), Pearl (1913, 1914a, 1914b), Jennings (1916), Wentworth and Remick (1916), Robbins (1917, 1918a, 1918b) and other workers in that field about that time. Lerner (1958, p. 3) defines that population genetics "is a discipline supplemented by, and overlaps in different ways, other compartments of the science of heredity and variation, which are designated as mathematical genetics and biometrical genetics." Through the 1920's and the early thirties considerable work was published by three of the main architects of population genetics: Fisher, Wright, and Haldane.

W. Johanssen of Denmark is credited for the distinction made of genotype and phenotype, distinguishing clearly environmentally caused variations which are not inherited from genetically caused variations which are inherited. Considerable work has been published on several of the factors that cause variations in milk production. Some of these are Eckles (1915), Hammond and Sanders (1923), Turner et al. (1924), Wylie (1925), Fohrman (1926), Sanders (1927, 1928), Gaines and Palfrey (1931), Gifford and Crosby (1933), Cannon (1933), and Copeland (1934). These early papers give some evidence of the effect of many variables on milk production and some rough measure of their importance. Some of these early workers offered rough methods for correcting records for a few of the environmental factors that cause differences in milk production. Since then, attempts have been made to evaluate quantitatively the roles played by environment and heredity.

Gowen (1934) used 14,000 Jersey Registry of Merit records to get his estimates for the roles of environment and heredity. From correlation coefficients between relatives such as paternal half sisters, maternal half sisters, full sisters, dam-daughter; and a measure of assortive mating from correlation between paternal grand dam and maternal grand dam, dam and paternal grand dam, he got estimates of hereditary and environmental variation under three different hypotheses. From this he concluded that 50 to 70 per cent of variation in milk production is accounted for by differences

in heredity, 5 to 10 per cent by differences in environment and the rest by dominance, assortive mating and environment common only to the cow. In this investigation, however, the calculations were made on the assumption that there was no correlation between parent and offspring due to environmental causes, which assumption cannot be justified.

Plum (1934, 1935) studied causes of differences in butterfat production of cows in Iowa Cow Testing Association records. By the analysis of variance he estimated the importance of different sources of variation in fat production. He concluded that it is likely that variance due to gene differences and gene combinations is somewhere between 20 and 40 per cent of total intra-herd variance.

Johansson and Hansson (1940), in their classical work with Swedish Red and White cattle, dealt in great length with the various sources that cause variation in milk production. Using the technique of correlation between relatives they estimated that 30 to 40 per cent was the hereditary part of total variance in yield of milk or butterfat.

Considerable work on the subject has been reported in recent years. Relevant portions of many of these will be reported and commented upon in the appropriate section dealing with the discussion of results obtained in the present investigation. This historical review merely refers to important early works on the subject and points out that within a

period of ten years in the early study of the subject conflicting estimates of the relative importance of heredity and environment had already been obtained.

## DATA

## Description

The data for this investigation are from 15 state-owned herds in Iowa. These herds are distributed in different parts of the state, as shown in Figure 1. Most of these herds are attached to welfare or penal institutions and are operated in part with labor from these institutions. The management and breeding policy are under the supervision of a central administration which often consults with the Dairy Husbandry staff at the Iowa State University of Science and Technology. One herd is the University herd under the direct control of the Animal Husbandry Department. These herds, by their location, represent different soil and climatic conditions of the state, although Iowa is in these respects more uniform than most states.

All the herds have Holstein-Friesian cattle, the University herd having other breeds in addition. Only the records of the Holstein-Friesians were used in this investigation. The herds are in the Herd Improvement Registry testing program and all cattle included in this investigation are registered with the Holstein-Friesian Association of America. Information on each animal born or brought into the herd is maintained on I.B.M. punched cards at Iowa State. The records cover the period from 1940 to date. Each



lactation record for each cow is on a separate card. On each card there is provision for the following information:

Herd number

Cow number

Dam number

Sire number

Birth - year and month

Lactation number

Start of lactation - year and month

Age at start of lactation - year and month

Days in milk

Times milked - twice or thrice

Milk yield - in pounds

Per cent fat

Fat yield - in pounds

Description of record - Complete or incomplete

305 day, M.E., 2 X, milk yield - in pounds

305 day, M.E., 2 X, fat yield - in pounds

305 day, M.E., 2 X, 3.5 per cent milk yield - in  
100 pounds units

M-H - deviation of record from the herd average  
for the year

Except for earlier years, when many of the cards did not have information on birth and age at start of lactation, the punched cards had complete information.

The yield corrected for age, length of lactation and frequency of milking was thus available from the cards. Age corrections had been made according to the factors of Kendrick (1955). The records had been adjusted to 305-day lactation by discarding production beyond 305 days. For cows that dried off before the completion of 305 days from the start of lactation, production to the date of drying was used as their 305-day production. Production records of cows that died or were sold during the milking period before completing 305 days from the start of lactation were designated as incomplete records. Most of the records were made on two times milking. Those made on milking three times daily had been adjusted to twice-a-day milking. In H.I.K. testing program the milk from each cow for two consecutive milkings, night and morning, once a month is weighed and a sample is tested for fat content. The recorded lactation production is the sum of each test production multiplied by the number of days the test represents.

The purpose of this investigation is to estimate the size of sources of variation in butterfat production. It was decided to use production adjusted to 305-day, mature equivalent, twice-milking basis. A duplicate set of I.B.M. cards containing the relevant information was reproduced. Incomplete records were separated and discarded. Three of the herds did not have information available for the full period

from 1940 to 1956 and, hence, their records were not used.

### Tests

From the twelve herds 12,623 complete records from 4,487 cows were available. This averages 2.8 records per cow. The data were subjected to preliminary tests for normality of the distribution of the variates and to homogeneity of variances among the twelve herds.

The 12,623 records ranged from 23 pounds to 879 pounds and averaged 410 pounds of fat. Table 1 gives the frequency distribution of these records and Figure 2 shows this graphically. In Figure 3 the cumulative distribution is plotted to normal probability scale. This should be a straight line if the distribution was perfectly "normal." Visually there is no marked deviation from normality except at the two ends. The discrepancies are all in the lower 4 per cent (at the bottom of Figure 3) and in the upper 1 per cent (at the extreme top). The frequency distribution was tested for deviation from normality both for skewness and kurtosis according to Snedecor (1956, p. 201). The values obtained for  $g_1$  and  $g_2$  and their standard errors are  $-0.1359 \pm 0.0128$  and  $0.425 \pm 0.0436$ , respectively. The "t" values are 6.23 and 9.75 both larger than the "t" value at  $p:0.01$ . Certainly the distribution is not perfectly normal. Some skewness and kurtosis exist. The small positive value of  $g_2$  suggests a slight excess of records near the mean and near the

Table 1. Frequency distribution of fat records

Class interval	Frequency	Frequency as per cent of total	Cumulative frequency as per cent of total
20- 49	2	0.016	0.016
50- 79	8	0.063	0.079
80-109	18	0.143	0.222
110-139	28	0.222	0.444
140-169	55	0.436	0.880
170-199	80	0.634	1.514
200-229	149	1.180	2.694
230-259	270	2.139	4.833
260-289	467	3.699	8.532
290-319	816	6.464	14.996
320-349	1133	8.975	23.971
350-379	1437	11.384	35.355
380-409	1734	13.737	49.092
410-439	1712	13.563	62.655
440-469	1511	11.970	74.625
470-499	1281	10.148	84.773
500-529	888	7.035	91.808
530-559	552	4.373	96.181
560-589	241	1.909	98.090
590-619	135	1.069	99.159
620-649	52	0.412	99.571
650-679	25	0.198	99.769
680-709	13	0.103	99.872
710-739	6	0.048	99.920
740-769	2	0.016	99.936
770-799	2	0.016	99.952
800-829	4	0.032	99.984
830-859	1	0.008	99.992
860-889	<u>1</u>	0.008	100.000
Total	12623		

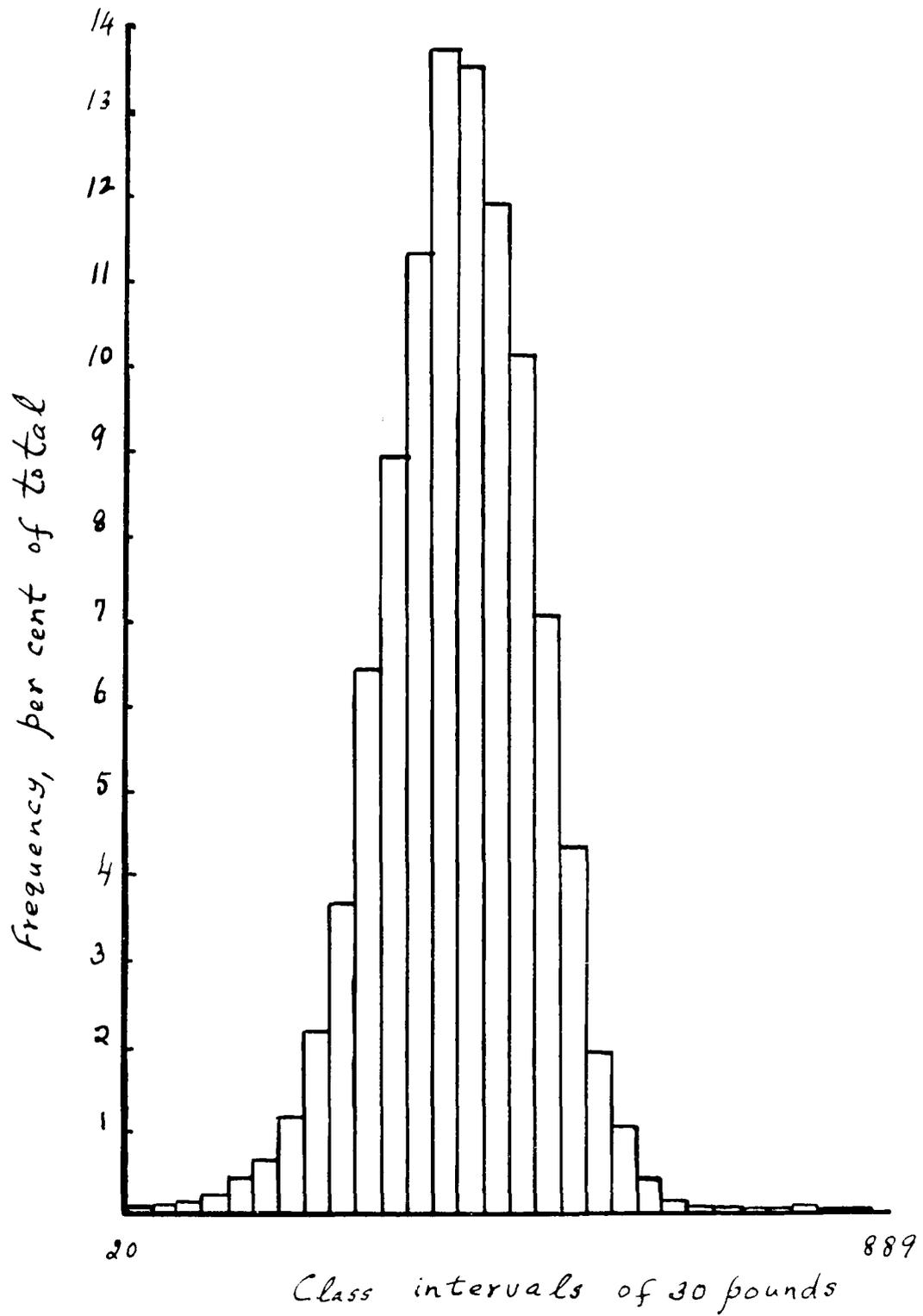
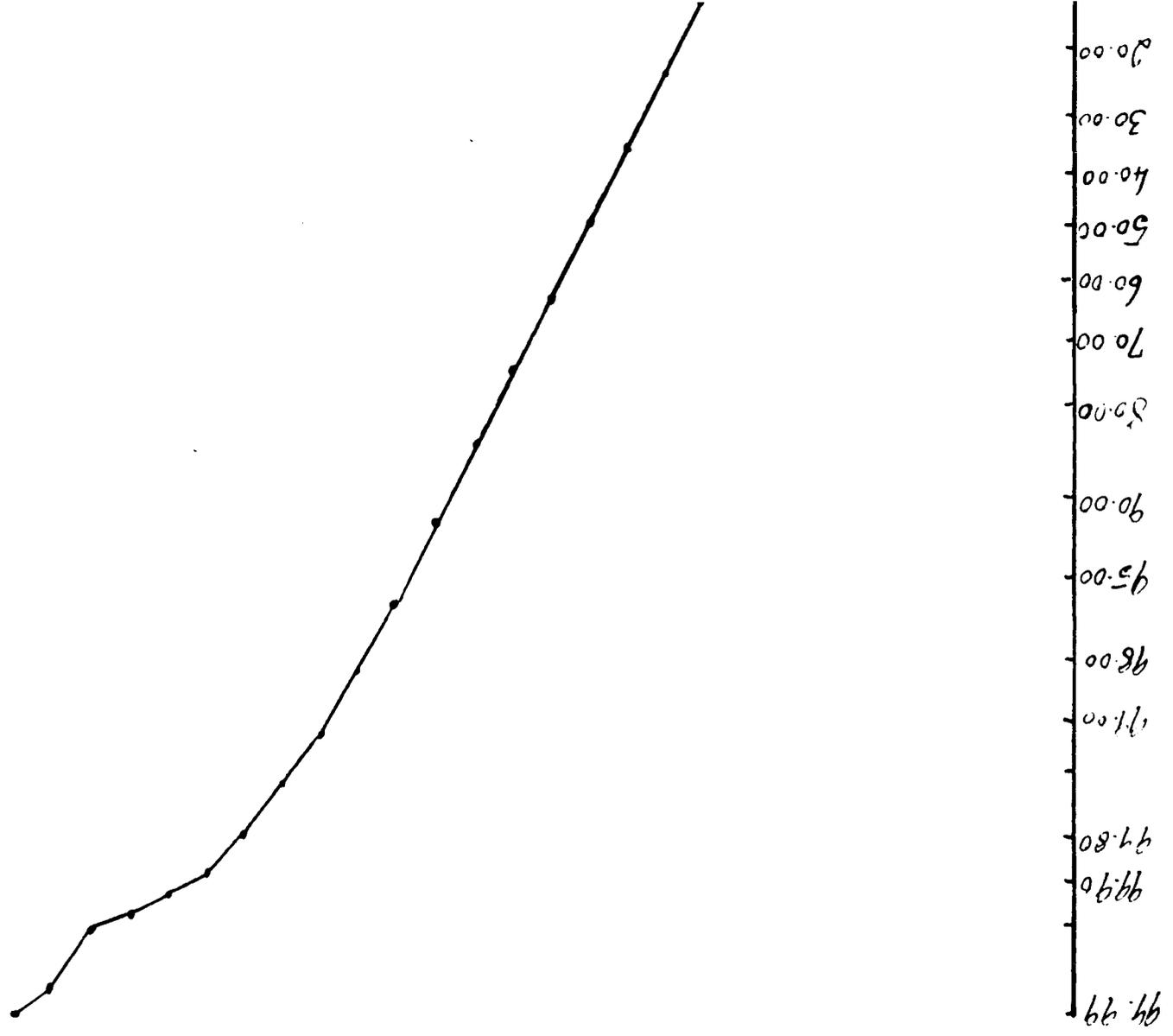


Figure 2. Frequency distribution of records.

relative frequency, percentage of total



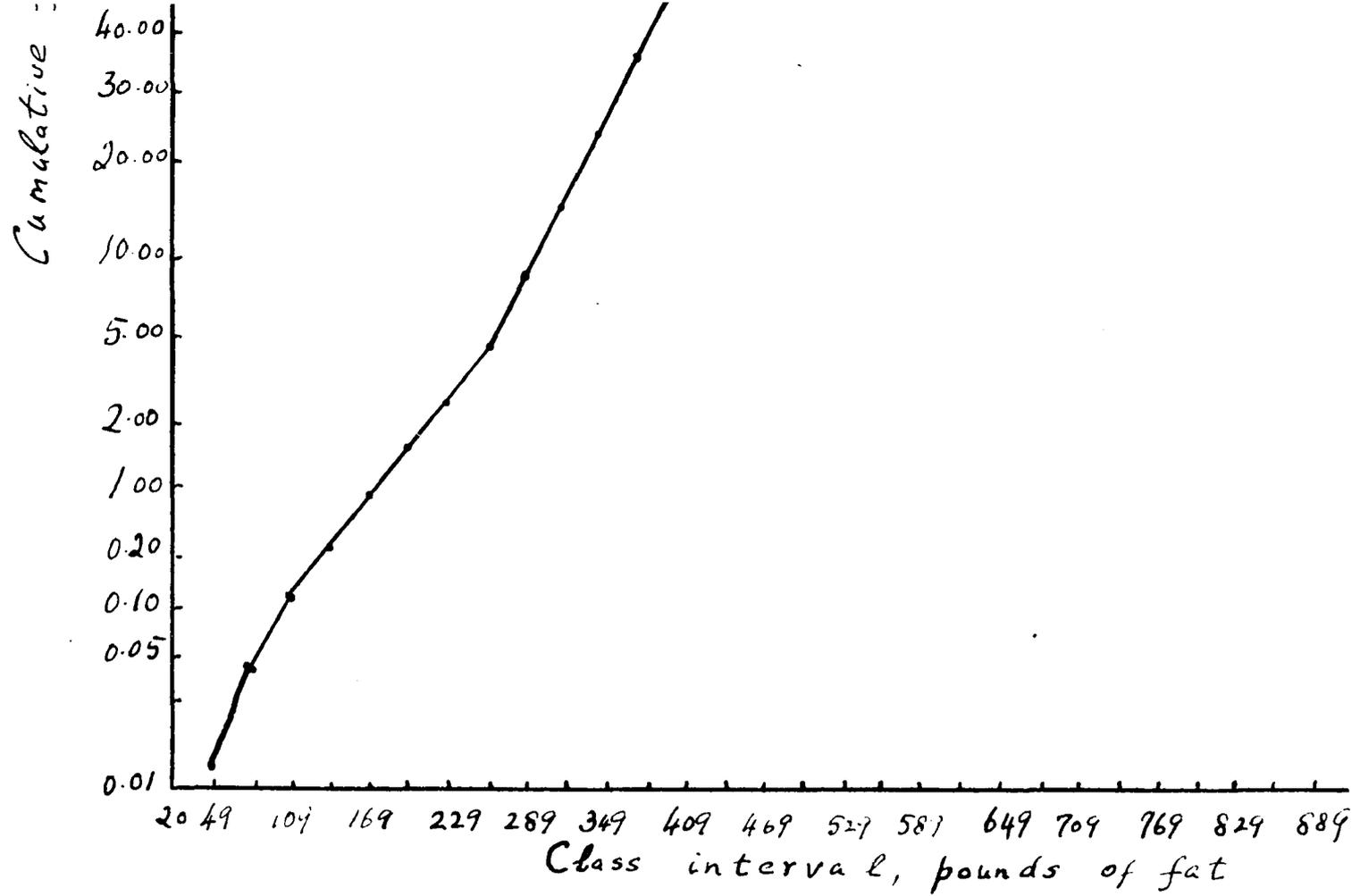


Figure 3. Cumulative frequency distribution of records to probability scale.

ends, with a corresponding depletion in the flanks. The small negative  $g_1$  indicates a slight asymmetry with an excess of records larger than the mean. This has drawn the peak of the frequency curve towards the right. However, this test is an approximate one and such large number of observations make the standard errors of the  $g$  values necessarily very small. These are very sensitive tests for picking up even small deviations from normality. The departures from normality are moderately small, although apparently not entirely chance ones.

The variances within herds differed widely among the twelve herds, as is shown in Table 2. The standard deviations ranged from 67.8 to 103.5 pounds and the coefficients of variation ranged from 16.0 to 26.9 per cent. In Figure 4 is plotted the relationship between mean and coefficient of variation. The scatter diagram does not indicate any definite relationship. According to Johansson (1953) low producing herds have smaller variation than high producing herds. That some herds in the investigation have fairly similar means but such widely different variances calls for probing the causes of this situation. If they could be investigated closely, the management practices at the different farms might suggest some explanation. For instance, if the cows at some farms are not fed according to their production but are given the same amount of feed, the potentially high

Table 2. Means and variances of fat production by individual herds

Herd no.	Mean fat lbs.	Variance (lbs.) <sup>2</sup>	Standard deviation lbs.	Coefficient of variation
1	431	6418	80.1	18.6
2	395	5759	75.9	19.2
3	413	10720	103.5	25.1
4	369	7430	86.2	23.4
5	350	8883	94.2	26.9
6	421	9309	96.5	22.9
7	424	4602	67.8	16.0
8	405	7630	87.4	21.6
9	415	9361	96.8	23.3
11	431	6428	80.2	18.6
12	425	6461	80.4	18.9
15	418	8012	89.5	21.4

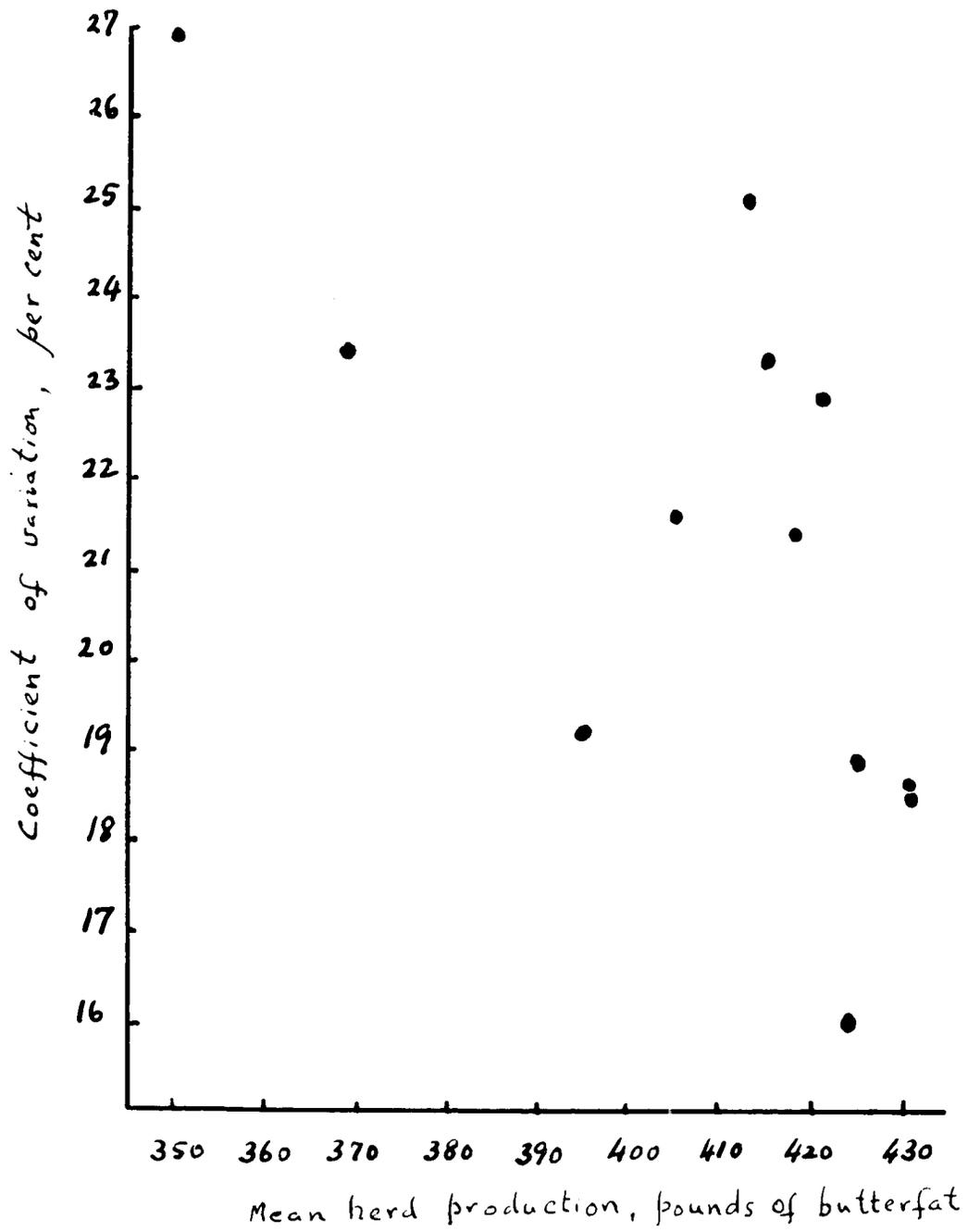


Figure 4. Mean and coefficient of variation relationship.

producing cows would not perform at their best. This practice could reduce variance. The opposite practice of favoring the high producers at the cost of average producers, in an attempt to get a few high records, could increase variation. Some of these herds seem to have grown in numbers while others have remained stationary. This might have affected their variances. The above views are offered, although without substantiation, in the hope of suggesting some investigation planned to find why herds such as 3, 5, 6 and 9 differ so widely in variation.

Bartlett's test for homogeneity of variances, according to Anderson and Bancroft (1952, pp. 141-142), yielded  $Q$  and  $l$  values of 131.04 and 1.0004, giving  $Q/l$  equal to 130.98. The  $\chi^2$  value for eleven degrees of freedom at  $p:001$  is 24.72. Clearly the different herds do not have similar variances.

## METHOD OF INVESTIGATION

With the widespread use of bulls by artificial insemination in the last decade, the question of environment confusing the classification of bulls according to their real merit has drawn the attention of workers in dairy cattle breeding. McBride (1958) in a review paper reports on different approaches of prominent workers to this problem. Environmental differences of concern are classified as macro-environmental and micro-environmental. Examples of different macro-environmental influences are different climatic conditions, differences in management from herd to herd, etc. Micro-environmental influences are those fluctuations which occur when all animals are apparently treated alike.

One can easily see that sources such as herd, year and season in which records are made can explain some of the differences between records. Other sources could be individual differences between cows due to their genotypes, permanent environmental effects brought about by disease or injury, the number of records prior to the current one, the previous calving interval, the previous dry period, the current calving interval, the current dry period, frequency of milking, age at which the record is made, etc.

Differences between herds could be due entirely to differences in management or partly to those and partly to

differences from herd to herd in average genetic merit of the cattle. Managerial differences could arise from differences in quality of feed available, the differences in climate or weather among locations of the herds, differences in labor available, supervision of the herd, health of cattle, etc.

The comparatively long generation interval in cattle usually spreads any genetic improvement over a period of several years. During these years the overall feeding and management of cattle may also tend to advance. However, periodic or irregular changes within this overall trend are often noticed. These could be caused by situations such as an unusually long winter or a hot summer resulting in lowered crop production, or by temporary economic upsets involving periods of recession or prosperity which could cause excess or shortage of feed or labor. These short term changes can cause variation in records. Year to year differences could be expected to show the effects of these sources of variation.

A source of variation considered important, since early days, is the period of the year the record is made. During the first few weeks of the lactation a cow gives a greater amount of milk for a given length of time than for a similar length at a later stage of the lactation. Obviously the first few weeks production would influence the whole lactation more than the production during an equally long period

in the latter part of the lactation. If an unfavorable environment at the early stage of lactation reduces production, it will have more influence on total production than if the unfavorable environment occurred later. However, this factor may not be very important in an area where seasonal changes are not severe or where steps are taken to alleviate the effects of these changes. In studying this source of variation, choosing the criterion to designate the different periods is difficult. If one examines first the level of production at different times within the year and then chooses the periods which contrast most, a definite bias is introduced. While this could be justified if the purpose is only to measure the extent of this source of variation in the sample studied, it will not be fair to accept these estimates as unbiased estimates of the population the sample represents. An obvious solution is to divide the year into calendar months. If total records available for a study are not sufficiently large, this would result in small numbers in each group, and the group averages might not reflect the real effect for the particular month. The argument for studying and correcting for a seasonal effect is that changes in weather and in feed supply are causes of differences in production. Monthly classification may be too fine a subdivision of the year to show clearly what are really seasonal effects. This suggests dividing the year

into periods on the basis of meteorologically distinct seasons. While this is satisfactory from the point of view that seasonal changes cause changes in production, the question remains as to the best criterion for classifying a record in one season or another. The most logical basis seems to be to classify records according to the date at which the record started. This does, in many instances, place a record in one season, while the high producing stage of the lactation really occurs in the following season. This would happen, for instance, while classifying as a summer record a record started in late summer, although most of the heavy production of the lactation will be made in the following fall.

The average production of all records combined per month was calculated. The following classification, similar to the usual meteorological division of seasons, was made:

Winter - Jan. 1 to March 31

Spring - April 1 to June 30

Summer - July 1 to Sept. 30

Fall - Oct. 1 to Dec. 31

Since the spring and summer production averaged about the same and the fall exceeded the winter by only a little (Table 11), the four seasons were classified into two major seasonal periods; spring and summer as the warm period with much pasture, and fall and winter as the cool period with

much of the feed fed indoors as harvested and processed crops. The seasonal effect was studied by the above three methods of classifying according to date of start of lactation: (1) by month, (2) by season and (3) by half-year period.

A major purpose of the investigation was to study main environmental effects and the interactions among them by analysis of variance and to get components for the different sources. Including more factors will make the analysis more unwieldy and will decrease the number of observations in the smaller subgroups. Therefore, all the other minor factors that could cause differences in production, such as calving interval, dry period, lactation number, etc., were ignored. Some evidence from an earlier study, Plum (1935), indicates that length of dry period is of very minor importance with D.H.I.A. records in Iowa. He estimated that one per cent of the total variance was caused by differences in length of dry period.

The effect of age of the cow on production is well known to be important, especially as between younger ages. Including that as a separate source of variation would further complicate the analysis. The records available were therefore adjusted first for age. While age corrections may not be perfect, there is evidence, Lush and Shrode (1950), that the most commonly used age correction factors will remove

most of the variance caused by age.

Another major purpose of the investigation was to estimate the genetic components of variation. To have these components free of environmental effects the analysis must be done within common environmental groupings. With plant and small animal data this may be approached closely, but with cattle data this is difficult. With records extending over a seventeen-year period and with two to twelve seasonal groupings within each year it is not possible to classify a sufficient number of paternal and maternal half sisters within groups which have a common environmental classification. Therefore, it was decided first to estimate the magnitude of the environmental sources of variation and then on the basis of information obtained, to correct the records for the major environmental effects. The genetic analysis was then made on the corrected records. The genetic analysis was done in three different ways: (1) as an hierarchal classification - by sires and by dams within sires, (2) by using full sister records only, and (3) as a cross classification by sires and by dams.

All available records were used to get estimates of the effects of the three major environmental factors studied and of the interactions among them. Of these records, a little more than two-thirds were second and later records of a cow. Since in these herds a criterion of culling could

have been the performance of a cow in her first lactation, second and later records could have been made by cows that had survived culling. This could lead to some confounding of selection effects with year effects and possibly even with season effects. In one year there could be a certain proportion of first records while in the next year there could be a highly different proportion of first records, and part of the year difference could be due to difference between records because of selection. To avoid this source of confusion, it was decided to get another set of estimates of these environmental factors using only the first lactation records. Use of first lactation records does not remove time trends from the year effects. A first lactation record of an average cow in a particular year could be considerably different from a first lactation record of an average cow ten years later, due either to change of genetic merit of cattle or to change in average environment during these ten years or to both. To free the year effects from most of any steady time trend, the year component was also obtained by conducting the analysis with the years classified into four periods, each of 4-5 consecutive years and getting year components within these periods.

## RESULTS

## Analysis of Variance Technique

The technique of analysis of variance was introduced by Fisher in 1925 in the first edition of his book "Statistical Methods for Research Workers." In early days, this technique was mostly used in testing significance of differences between group means in planned experiments. During recent years, considerable attention has been given to the various assumptions, which must hold true if the results of using this technique are to be valid and to the various applications of this method to biological problems encountered either in designed experiments or in data collected from the field.

Crump (1946), in an account of estimating variance components, describes the two major uses of analysis of variance with a model:

$$y_{hij} = u + a_h + b_i + ab_{hi} + e_{hij}$$

In the first instance, under the assumptions that  $e_{hij}$ 's are normally and independently distributed with zero mean and constant variance and  $u$ ,  $a$ ,  $b$  and  $ab$  are parameters of fixed effects, the problem is to estimate  $a_h$ ,  $b_i$ ,  $ab_{hi}$  and  $u$ , and to test the null hypothesis that the parameters in any set are all equal to zero. In the second instance, under the only assumption needed that  $a_h$ ,  $b_i$ ,  $ab_{hi}$  and  $e_{hij}$  are random

variables independently distributed, the parameters estimated are the variances of the random effects. It should be noted that here, in the second instance, no assumption is made about the form of the distribution. The components of variance are estimated by equating the expectations of the mean squares to the mean squares observed in the analysis and solving for the components. If the random elements in the model do follow a normal distribution, the confidence limits for the components could be computed.

Crump's work was followed by Eisenhart (1947) who studied the assumptions underlying the analysis of variance. He also dealt with the two different functions of analysis of variance: (1) the estimation of fixed relations among means of subgroups and (2) detection and estimation of components of variation. However, he indicates that often no distinction is made, since the analysis of variance is used mostly in tests of significance where the computational procedure and tests are the same whether the desire is to infer either the existence of a fixed difference among true means of subgroups or the existence of a component of variation ascribable to a particular random factor. In problems of estimation with model I (fixed effects) the parameters are means, while in model II (random effects) the parameters are variances. According to him the necessary conditions for solving problems of model I are that effects are additive,

errors are normally and independently distributed around zero mean and have constant variance. Under model II, the assumptions needed are that the random effects are additive, independently distributed around zero mean and have constant variance. He also discusses the question of what is to be considered a fixed effect and what is to be considered a random effect. If the conclusions to be obtained from a study are to apply only to the finite population actually studied, the effect is fixed, while if they are to be applied to an infinite population, which the sample studied represents, the effect is random.

Cochran (1947) saw the need to specify consequences when assumptions for the analysis of variance are not satisfied. From several investigations, he concluded that non-normality of the distribution does not introduce serious error in tests of significance. Heterogeneity of error variance will lower efficiency in estimating treatment effects. Correlation among errors affects the estimation of treatment means, but proper randomization tends to destroy the correlation. Effects of non-additivity are not important unless the departure from additivity is very serious. If the treatment effects do not exceed 20 per cent, additive relationship is likely to be a good approximation in most cases.

Henderson (1953) adapted the technique of getting components of variance to the non-orthogonal situation, when

different subgroups do not have equal numbers of observations. This is the situation encountered almost invariably with livestock data. He describes in detail the steps of getting components in three different situations: (a) Method 1, when all variables are considered random, (b) Method 2, when some variables are considered random and others fixed, and (c) Method 3, when variables in the model are correlated.

In the present investigation all variables are considered random. The environmental variables considered are herds, years and seasons. The variables involved in the genetic analysis are sires and dams. The purpose of the analysis is to estimate the variances these variables produce in a population which these observations represent; it is fair to assume that these variables are random.

The preliminary investigation showed that the variates depart from normality of distribution and homogeneity of variances. This and the non-orthogonal nature of the data make a bit uncertain the tests of significance. Yet the major purpose of this investigation was to estimate the components of variance ascribable to the variables studied and for this the failure of the above assumptions to be entirely valid is little, if any, hindrance.

## Findings on Environmental Factors

Components

For the analysis of variance from environmental factors the model assumed is:

$$y_{ijkl} = u + h_i + a_j + s_k + h_{aj} + h_{sik} + a_{sjk} + h_{asijk} + e_{ijkl}$$

where,

$y_{ijkl}$  is the  $l$ th observation of the  $k$ th season  
of the  $j$ th year of the  $i$ th herd

$u$  is the general mean

$h_i$  is an effect particular to the herd,  $i =$   
1 . . . 12

$a_j$  is an effect particular to the year,  
 $j = 1 . . . 17$

$s_k$  is an effect particular to the season,  
 $k = 1 . . . 2, \text{ or } 4 \text{ or } 12$

$h_{aj}$  is an interaction effect particular to  
the herd and year

$h_{sik}$  is an interaction effect particular to  
the herd and season

$a_{sjk}$  is an interaction effect particular to  
the year and season

$h_{asijk}$  is an interaction effect particular to  
the herd, year and season

$e_{ijkl}$  is a random effect peculiar to an observation

The assumptions made in the model are (1) additivity of main effects and interactions and (2) the effects are independently distributed. When the analysis was done with all

records, the same cow had two or more records in many cases and, since these records are related, their errors are not independently distributed. With the study restricted to first lactation records this difficulty does not arise.

Variances ascribable to these random effects and interactions are  $\sigma_h^2$ ,  $\sigma_a^2$ ,  $\sigma_s^2$ ,  $\sigma_{ha}^2$ ,  $\sigma_{hs}^2$ ,  $\sigma_{as}^2$ ,  $\sigma_{has}^2$ , and  $\sigma_e^2$ .

Components were obtained by Method I of Henderson (1953).

From the model one can develop expectations for sums of squares. This development of expectations is shown by Kempthorne (1957, pp. 238-243). Most of the calculations involved in getting sums of squares and their expectations were made on I.B.M. machines.

Table 3 gives the coefficients of the expectations for the uncorrected sums of squares for the first analysis, given in Table 4. From these, the expectations of the corrected sums of squares for the different effects and interactions were easily calculated. For instance, the expectation of corrected sum of squares for herds was obtained by subtracting from the figure for uncorrected sums of squares in Table 3, the figure for the correction factor in the same column. Now the expectations were equated to the respective observed sums of squares. These provided eight simultaneous equations with eight unknowns,  $\sigma_h^2$ ,  $\sigma_a^2$ ,  $\sigma_s^2$ ,  $\sigma_{ha}^2$ ,  $\sigma_{hs}^2$ ,  $\sigma_{as}^2$ ,  $\sigma_{has}^2$ , and  $\sigma_e^2$ , the variance components. Solutions of these equations provided the estimates of these components.

As indicated under an earlier section, the analysis was conducted in two ways: (1) with all records and (2) with only the first lactation records. In the analysis with all records season was classified in three ways, according to when the lactation started: by month, by season and by half-year period. In the analysis of first lactation records season was classified only two ways: by season and by half-year period. The analyses of variance by these different methods are given in Tables 4 to 6.

The last equation shows  $\sigma_e^2$  directly. Substituting this in the other equations reduces the simultaneous equations to seven with that many unknown components. These equations were solved by using the I.B.M. 650.

The components of variance obtained are given in Table 7. These values expressed as per cent of the total variance are given in Table 8. The major differences between all records and first records are that season is more important and year is a bit less important in the former.

As mentioned earlier, the year component could include effects both of genetic improvement of cattle and of changes in environment. In an attempt to separate most of the effects of steady time trends from the effects of irregular year-to-year changes, the entire study was broken to four periods of four to five consecutive years. Such a period is approximately equal to the generation interval in dairy

Table 3. Coefficients in the expectations of uncorrected sums of squares<sup>a</sup>

Source	U	V(H)	V(A)	V(S)	V(HA)	V(HS)	V(AS)	V(HAS)	V(E)
Total	12623	12623	12623	12623	12623	12623	12623	12623	12623
Herd	12623	12623	773	1092	773	1092	83	83	12
Year	12623	1181	12623	1093	1181	126	1093	126	17
Season	12623	1181	774	12623	89	1181	774	89	12
Herd-year	12623	12623	12623	1370	12623	1370	1370	1370	204
Herd-season	12623	12623	965	12623	965	12623	965	965	144
Year-season	12623	1458	12623	12623	1458	1458	12623	1458	204
Herd-year-season	12623	12623	12623	12623	12623	12623	12623	12623	2328
Correction factor	12623	1154	754	1066	70	100	65	7	1

<sup>a</sup>All records; classified by month.

Table 4. Analysis of variance: seasonal classification by month of freshening<sup>a</sup>

Source	d/f	Sums of squares	Mean square
Herd (H)	11	5,655,038	514,094
Year (A)	16	1,441,919	90,120
Season (S) (month)	11	1,904,340	173,122
H x A	176	10,951,658	62,225
H x S	121	1,572,419	12,995
A x S	176	1,869,293	10,621
H x A x S	1,816	11,706,298	6,446
Remainder	<u>10,295</u>	<u>66,544,835</u>	6,464
Total	12,622	101,645,800	

<sup>a</sup>All records.

Table 5. Analysis of variance: classification by three-month season of freshening

Source	All records			First records		
	d/f	S.S.	M.S.	d/f	S.S.	M.S.
Herd (H)	11	5,655,038	514,094	11	1,920,628	174,602
Year (A)	16	1,441,919	90,120	16	746,415	46,651
Seasons (S)	3	1,232,644	410,881	3	100,056	33,352
H x A	176	10,951,658	62,225	176	3,655,915	20,772
H x S	33	549,628	16,655	33	467,579	14,169
A x S	48	734,985	15,312	48	526,399	10,967
H x A x S	522	4,638,130	8,885	443	2,543,798	5,742
Remainder	<u>11,813</u>	<u>76,441,798</u>	6,471	<u>3232</u>	<u>19,467,301</u>	6,023
Total	12,622	10,645,800		3962	29,428,091	

Table 6. Analysis of variance: seasonal classification by freshening date into half-year periods

Source	All records			First records		
	d/f	S.S.	M.S.	d/f	S.S.	M.S.
Herd (H)	11	5,655,038	514,094	11	1,920,628	174,602
Year (A)	16	1,441,919	90,120	16	746,415	46,651
Seasons (S)	1	979,909	979,909	1	76,442	76,442
H x A	176	10,951,658	62,225	176	3,655,915	20,772
H x S	11	307,502	27,955	11	165,067	15,006
A x S	16	367,620	22,976	16	202,830	12,677
H x A x S	176	1,077,784	6,124	168	817,971	4,869
Remainder	<u>12,215</u>	<u>80,864,370</u>	6,620	<u>3563</u>	<u>21,842,823</u>	6,130
Total	12,622	101,645,800		3962	29,428,091	

Table 7. Components of variance in absolute values

Source	All records			First records	
	Seasonal class.			Seasonal class.	
	Month	Season	Half-year	Season	Half-year
Herd (H)	426	426	412	443	439
Year (A)	24	23	7	85	70
Season (S)	149	123	149	15	27
H x A	903	856	907	750	820
H x S	57	20	38	73	49
A x S	28	14	38	21	39
H x A x S	62	197	4	74	-80
Remainder	<u>6,464</u>	<u>6,471</u>	<u>6,620</u>	<u>6,023</u>	<u>6,130</u>
Total	8,113	8,130	8,175	7,484	7,574

Table 8. Components of variance expressed as percentage of the total

Source	<u>All records</u>			<u>First records</u>	
	<u>Seasonal class.</u>			<u>Seasonal class.</u>	
	Month	Season	Half-year	Season	Half-year
Herd (H)	5.3	5.2	5.0	5.9	5.8
Year (A)	0.3	0.3	0.1	1.1	0.9
Season (S)	1.8	1.5	1.8	0.2	0.4
H x A	11.1	10.5	11.1	10.0	10.8
H x S	0.7	0.2	0.5	1.0	0.6
A x S	0.3	0.2	0.5	0.3	0.5
H x A x S	0.8	2.4	0.0	1.0	-
Remainder	79.7	79.6	81.0	80.5	80.9

cattle. This analysis was done only with first lactation records using both season and half-year classification of records. The sum of squares, due to differences among means of these groups of years, was removed from the sum of squares, due to difference among years. From this, the component, for the differences among years within these groups, was obtained. By either method of classification of record, this component, which measures the irregular year-to-year changes and little of the steady time trend that existed within the short period of four to five years, was zero. The difference between the component for year effects, obtained earlier (Table 8), and the above value (zero) was 1.1 and 0.9 per cent, respectively, for the two methods of classification of records. This component accounts for most of steady time trend measuring the genetic improvement and changes in environment.

#### Herd effect

The average lactation production per herd varied from 350 to 431 pounds, with the overall average production being 410 pounds. The analysis with first lactation records presented much the same picture as to the differences between herds, while the actual averages were slightly higher. While the range seems wide, most of this comes from unusually low production of two herds, numbers 4 and 5. Indeed, nine of

Table 9. Average production per herd

Herd No.	All records		First records	
	Number	Average	Number	Average
1	1082	431	395	440
2	845	395	211	390
3	1320	413	461	428
4	1466	369	420	380
5	364	350	138	356
6	575	421	177	417
7	870	424	207	425
8	1421	405	460	423
9	1252	415	427	435
10	1220	431	357	442
11	940	425	285	427
12	1268	418	425	424
Overall average	12623	410	3963	420

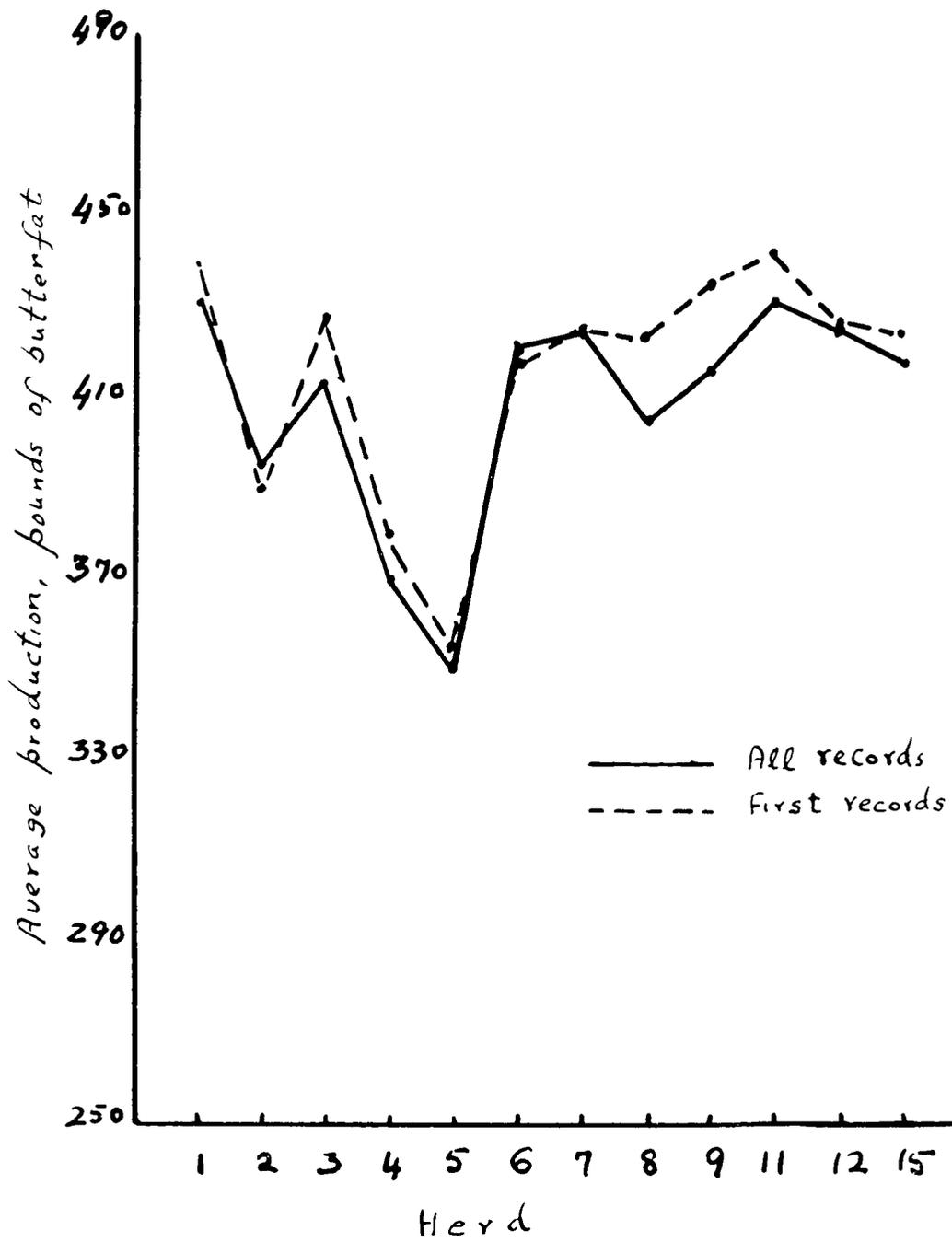


Figure 5. Production by herd.

the twelve herds have averages ranging only between 405 and 431. The first records show a similar picture except that only one herd, instead of two, had an average far below the others. The data suggest that a more detailed study into the circumstances surrounding especially herd number 5 but also herds 2 and 4 might establish some reasons for the differences between herd averages. Table 9 gives the average production of each herd with all records and with first records. Figure 5 shows this graphically.

Expressed as a component, herd differences account for 5 to 6 per cent of the total variance. This estimate is about the same when the analysis was done with all records as with first records. The herd-by-year interactions, however, account for nearly twice that much.

#### Year effect

The average lactation production per year varied from 389 to 432 pounds, with most years averaging between 400 to 420 pounds. The situation with first records was only slightly different, averages ranging from 388 to 443 pounds. Table 10 gives the average production in each year, both for all records and for first records only. Figure 6 has this information plotted. The differences between years are smaller than those between herds, as might be inferred from the importance of year in Table 8. However, the figure

Table 10. Average production per year

Year	All records		First records	
	Number	Average	Number	Average
1940	546	418	203	408
1941	639	419	201	418
1942	685	406	221	411
1943	654	398	180	388
1944	617	396	193	403
1945	697	389	190	408
1946	661	401	197	430
1947	761	402	291	412
1948	763	408	215	431
1949	847	409	259	434
1950	863	416	288	419
1951	790	400	219	412
1952	846	412	288	409
1953	776	419	205	434
1954	806	422	270	436
1955	817	413	239	424
1956	<u>855</u>	<u>432</u>	<u>304</u>	<u>443</u>
Overall average	12,623	410	3,963	420

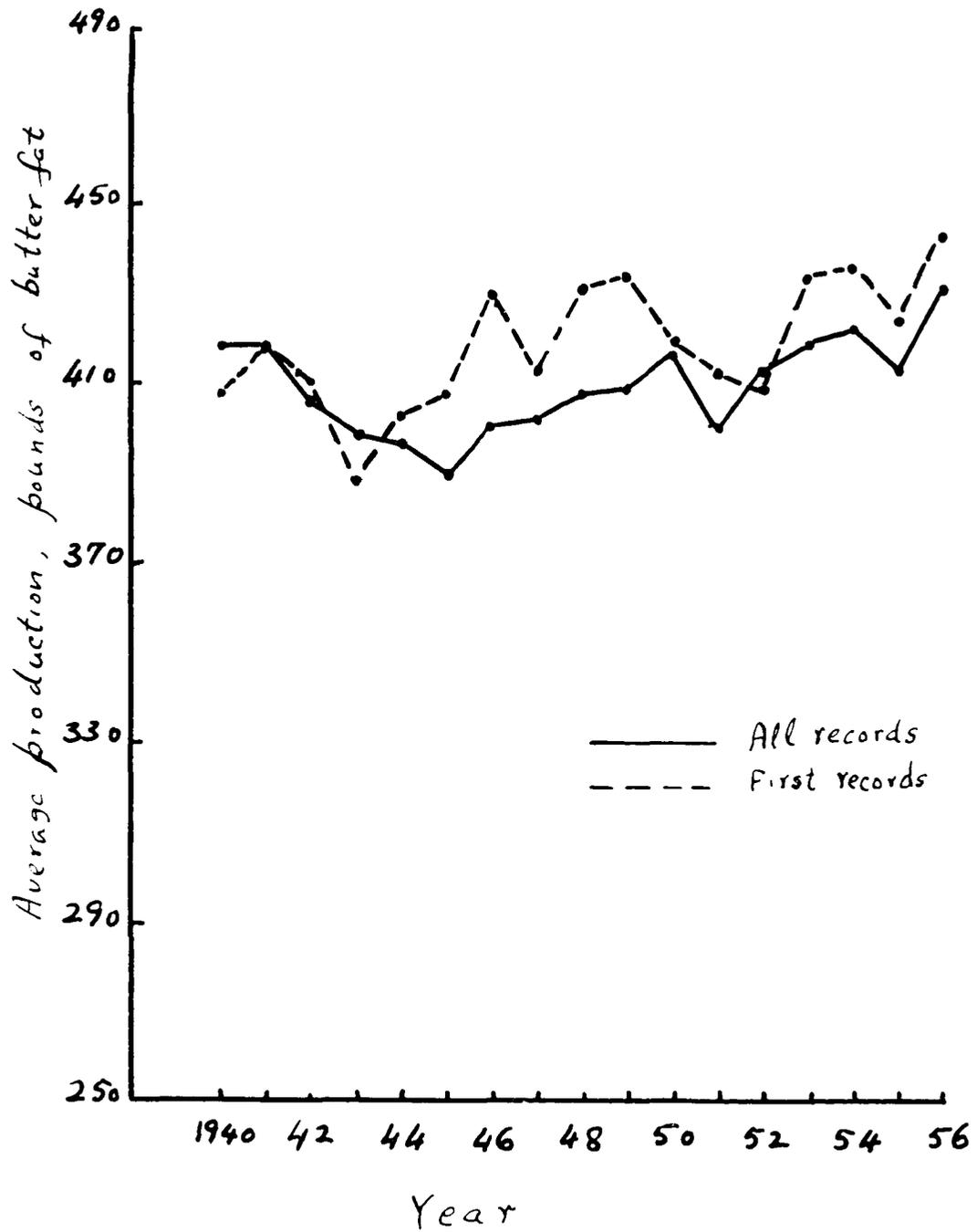


Figure 6. Production by year.

indicates a general upward trend in production during the period investigated. This would have seemed much more pronounced if the investigation had not begun until 1943 or 1944.

Expressed as a component, year differences account for only 0.3 per cent, or less, of the total variance when all records are used, and about 1.0 per cent when only first records are used.

### Season effect

Seasonal classification of start of lactation was made on three bases: (1) by month, (2) by season, and (3) by half-year period. The average productions for the different seasonal classifications with all records and with first records are given in Table 11. The trend is shown graphically in Figure 7. Monthly average production with all records varied from 390 to 427 pounds. The lower figures occurred in the spring and summer months. Appreciably fewer records were made during the spring months. Average production by three-month seasons of calving ranged from 400 to 424 pounds when all records were considered. The situation with first records was about the same, except that the averages were higher than for all records.

The warm period, in Table 11, is the pooling of spring and summer production while the cool period is the pooling

Table 11. Average production by months and by seasons

Class	Number of records	Average pounds	Average production	
			All records	First records
Jan.	1169	412		
Feb.	1096	416		
March	1042	407	Winter	412
April	833	407		
May	817	405		
June	942	393	Spring	401
July	1055	393		
Aug.	1167	390		
Sept.	1106	418	Summer	400
Oct.	1061	424		
Nov.	1172	427		
Dec.	1163	422	Fall	424
			Warm period	401
			Cool period	418

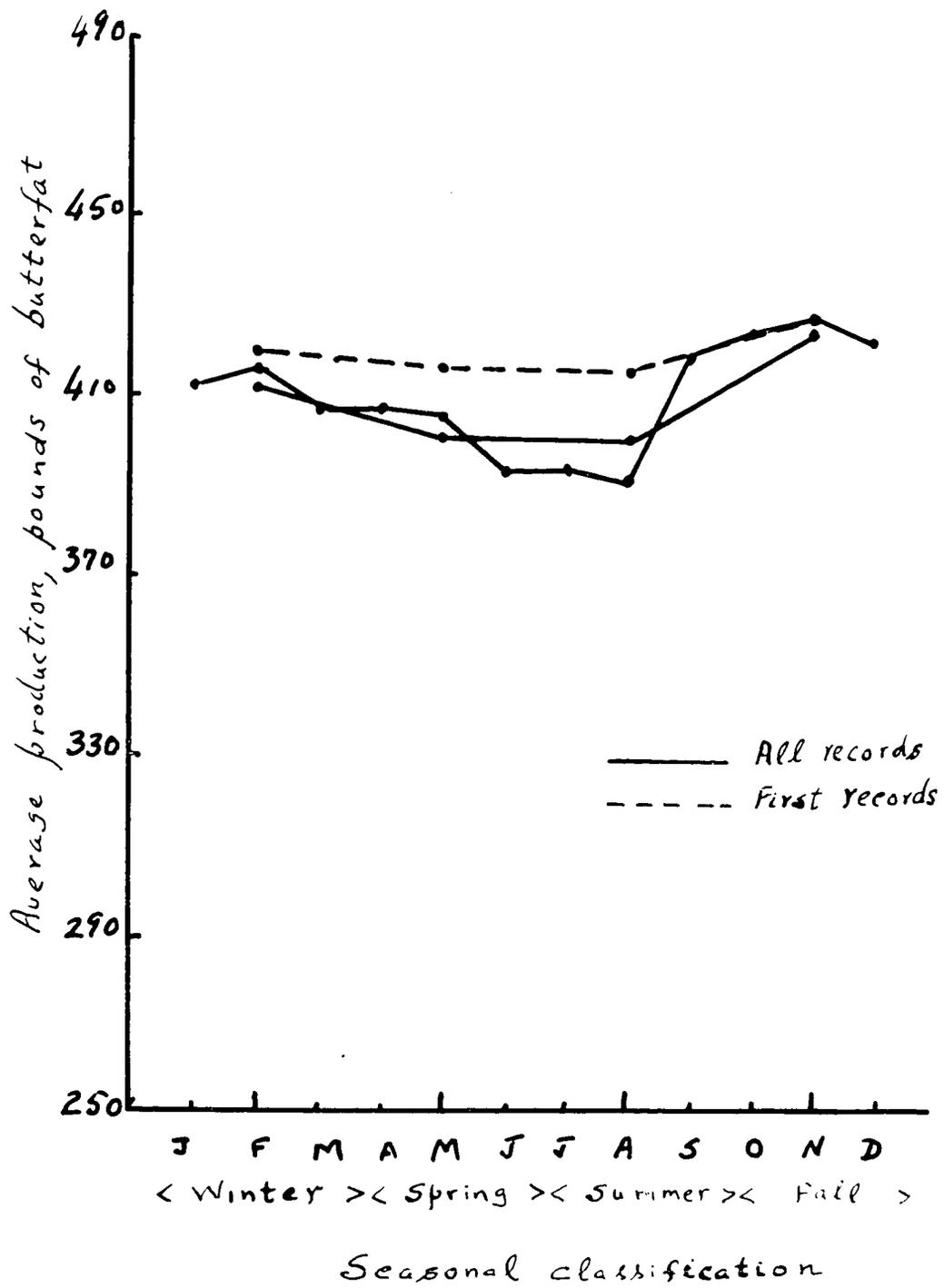


Figure 7. Production by time of year.

of fall and winter production. Whether the study was made with all records or was restricted to the first records, the cool period production was above the warm period production, the difference being 17 and 8 pounds, respectively, per lactation.

It is evident that records made at different times of year differed somewhat, although these differences were less extreme than were the differences between herds. Expressed as a component of variance, season accounted for 1.5 to 1.8 per cent of total variance of all records and from 0.2 to 0.4 per cent of the variance in first records. The smaller seasonal effect on first records is in contrast with the year effects which were larger on first records than on all records.

### Interactions

With the three factors studied: herd, year and season, three two-factor interactions and one three-factor interaction are possible. Interaction is a measure of the deviation of a subgroup average from the value expected for that subgroup from the information on the averages of the major groups to which the subgroups belong. For instance, in this investigation, if the herd average for a particular year is different, by more than a sampling error, from what one would expect from the average of that herd over all years and

average of that year over all herds, then there is interaction between herd and year for that particular herd-year subgroup.

Table 12 shows the average production for each year for each herd separately. The yearly changes for the different herds are plotted in Figure 8 for only four of the twelve herds, chosen at random, to avoid crowding too many lines in one figure. Obviously the production changes from one year to another are not the same in every herd. In statistical language the interaction between herds and years seems large. The component of variance analysis confirms this. The component for interactions between herd and year varied from 10.0 to 11.1 per cent of total variance and was the biggest among all the components in Table 8, except that for "remainder."

The extent of interaction between herd and three-month season can be seen from Table 13 and Figure 9. Figure 9 shows only eight of the twelve herds, chosen at random. Except for two of the twelve herds the trend of seasonal change is fairly common for the different herds. This indicates only small interactions between herd and season. The components obtained by the analysis confirm this. With the different classifications of season and using all records or only the first records, this component ranged from 0.2 to 1.0 per cent of the total variance.

Table 12. Average production per herd per year<sup>a</sup>

Year	Herd											
	1	2	3	4	5	6	7	8	9	11	12	15
1940	412	396	508	385	340	384	435	406	444	426	484	362
1941	375	412	507	377	301	366	467	409	439	429	471	428
1942	386	386	462	368	262	329	460	391	422	426	434	447
1943	408	383	427	354	290	332	475	353	429	451	397	425
1944	395	410	446	321	337	379	455	359	447	490	405	369
1945	411	365	445	314	415	366	437	355	388	414	441	361
1946	450	361	401	340	385	399	426	417	398	402	446	376
1947	464	351	412	351	395	371	399	425	399	406	423	376
1948	477	350	384	379	381	404	401	420	434	405	458	375
1949	441	348	428	370	374	488	410	416	399	432	402	400
1950	436	428	405	377	336	485	405	390	420	429	416	451
1951	445	405	378	367	328	438	384	418	382	443	402	397
1952	445	420	368	391	334	490	414	422	389	449	416	426
1953	429	440	380	400	312	512	416	429	399	440	385	470
1954	457	418	366	385	332	491	433	445	438	422	443	442
1955	447	412	367	382	392	463	417	414	401	433	413	459
1956	<u>457</u>	<u>435</u>	<u>423</u>	<u>379</u>	<u>480</u>	<u>445</u>	<u>413</u>	<u>415</u>	<u>442</u>	<u>465</u>	<u>425</u>	<u>463</u>
Herd average	431	395	413	369	350	421	424	405	415	431	425	418

<sup>a</sup>All records.

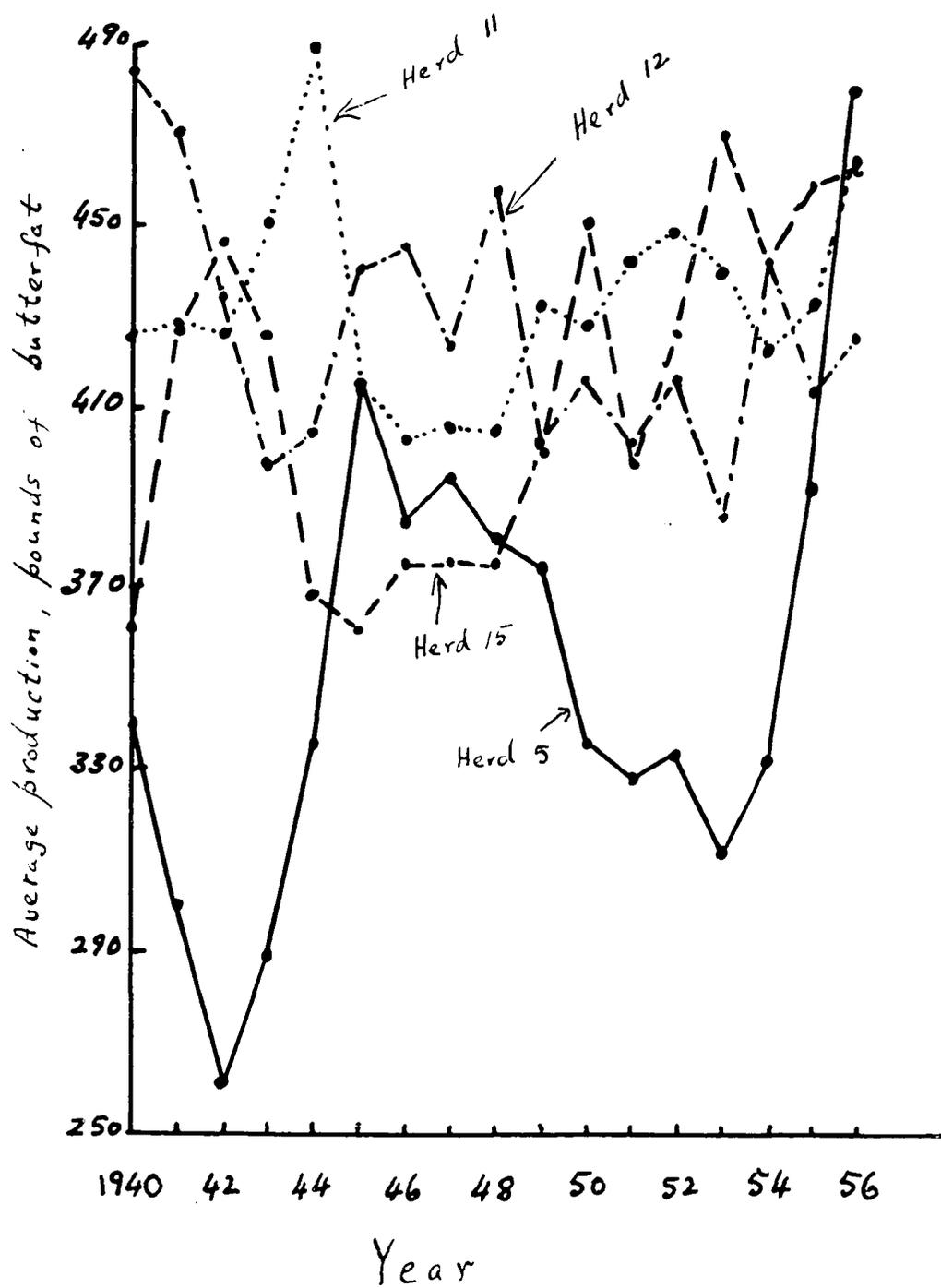


Figure 8. Herd by year interaction, illustrated with four herds.

Table 13. Average production per herd per season<sup>a</sup>

Herd	Season			
	Winter	Spring	Summer	Fall
1	432	424	420	444
2	384	392	393	407
3	426	404	392	431
4	371	365	360	379
5	360	334	336	361
6	416	430	411	432
7	440	416	404	435
8	413	395	392	417
9	418	396	406	436
11	438	417	421	446
12	415	412	429	438
15	<u>414</u>	<u>402</u>	<u>417</u>	<u>437</u>
Seasonal average	412	401	400	424

<sup>a</sup>All records.

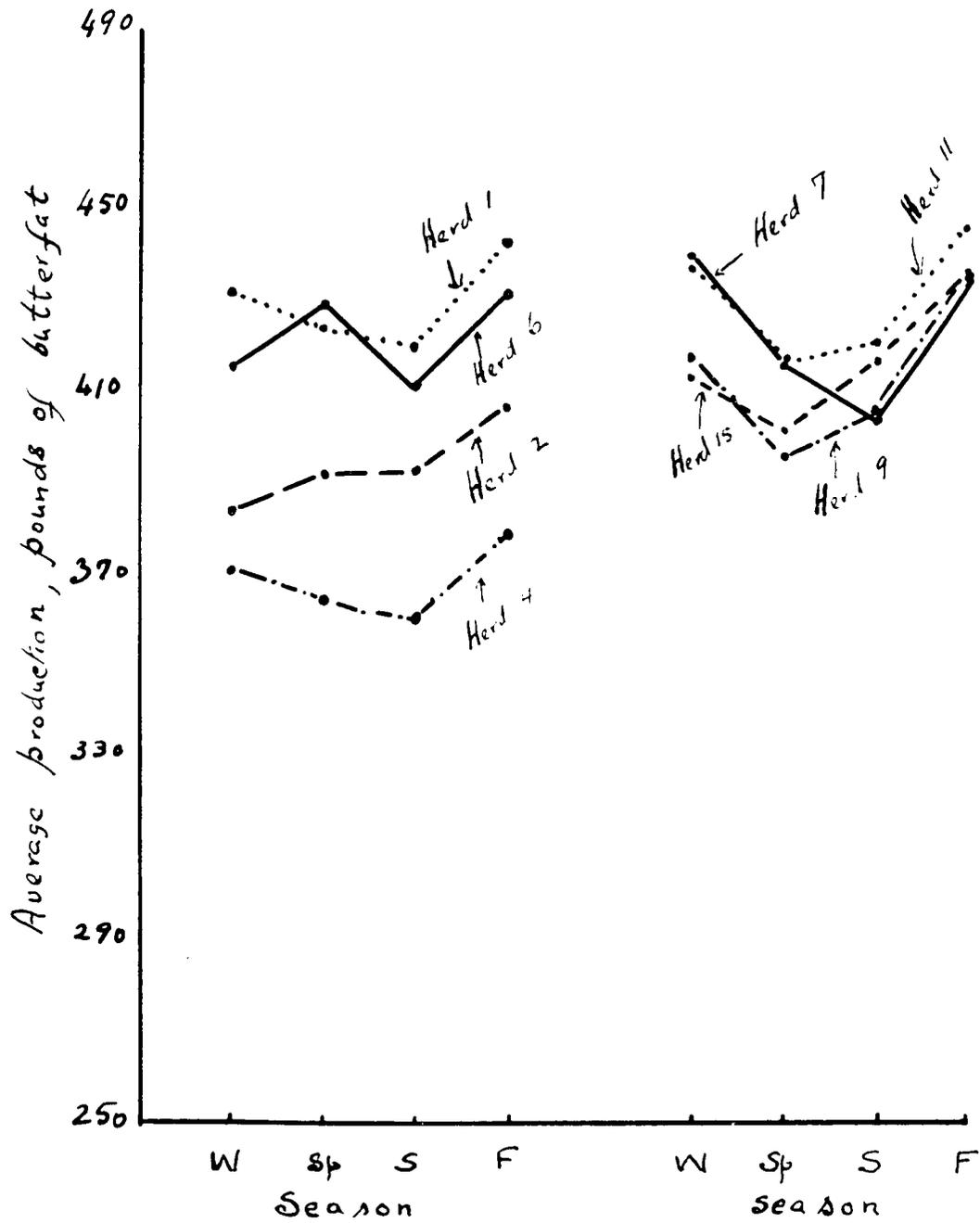


Figure 9. Herd by season interaction, illustrated with eight herds.

Table 14. Average production per year per season<sup>a</sup>

Year	Season			
	Winter	Spring	Summer	Fall
1940	410	410	415	432
1941	417	422	410	426
1942	419	410	385	410
1943	407	387	385	407
1944	405	398	381	399
1945	388	380	383	402
1946	393	390	400	415
1947	409	395	392	411
1948	398	398	396	437
1949	425	388	382	437
1950	413	409	412	430
1951	404	388	391	416
1952	406	402	401	434
1953	432	399	408	437
1954	421	415	418	434
1955	421	408	392	433
1956	<u>429</u>	<u>419</u>	<u>436</u>	<u>440</u>
Seasonal average	412	401	400	424

<sup>a</sup>All records.

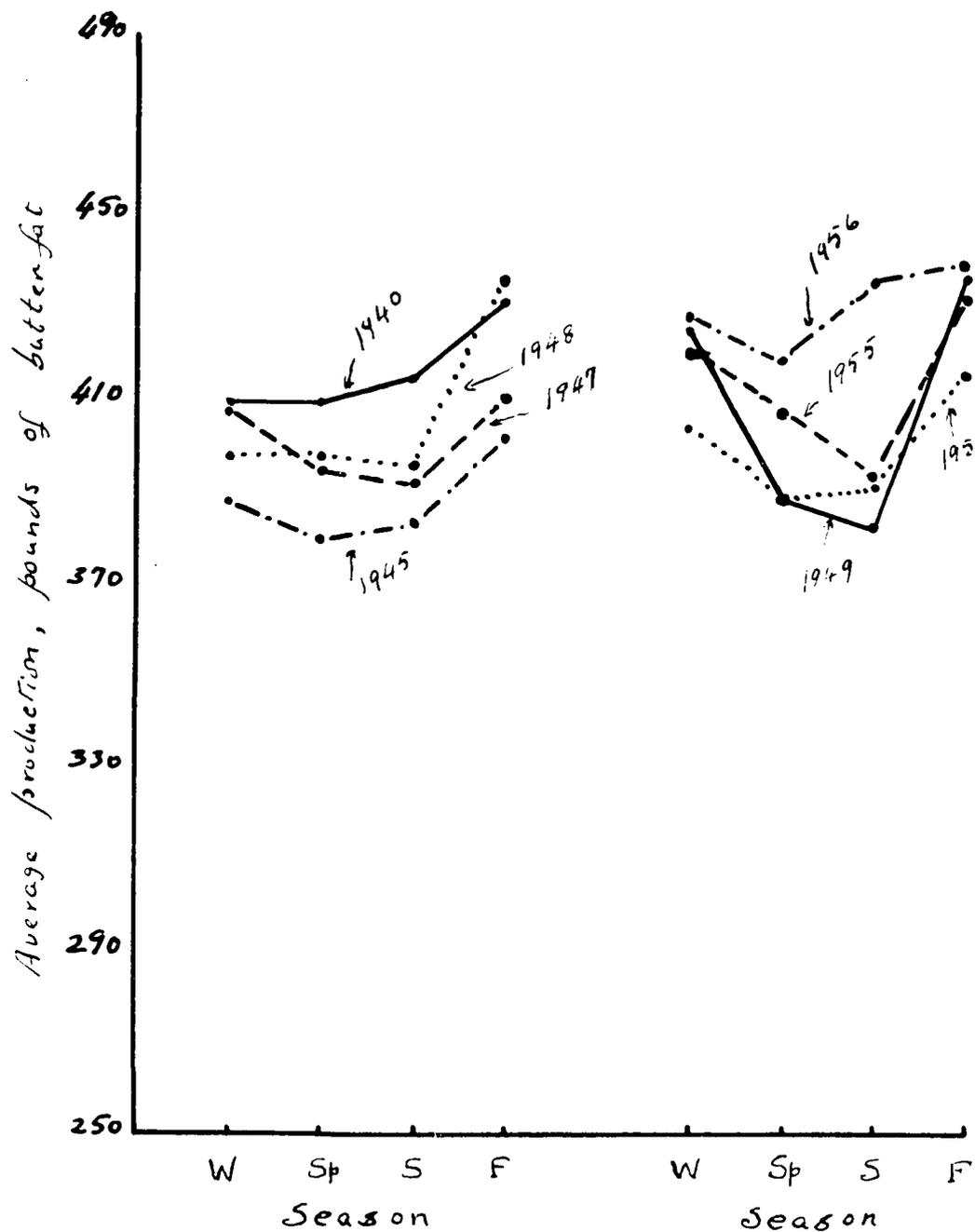


Figure 10. Year by season interaction, illustrated with eight years.

Similarly Table 14 and Figure 10 indicate that the interaction between year and season is negligible. The graph shows only eight of the seventeen years, chosen at random. The component obtained from the analysis (Table 8) varied from 0.2 to 0.5 per cent of the total variance, thus confirming the small importance of this interaction.

The three-factor interaction component varied from 0.0 to 2.4 per cent of total variance.

#### Findings on Genetic Factors

The findings reported in the preceding section show clearly that, among the environmental factors studied, differences between herds and the herd by year interactions are important sources of differences between records. The purpose of the study reported in the present section was to estimate the importance of genetic factors in causing differences between records. If records are influenced by environmental factors and if these factors are not randomly distributed in the genetic analysis, these factors may cause biased estimates of the genetic parameters. Therefore, it was necessary to remove from the records the effects of herd and herd by year interactions. This was done by taking every record as a deviation from the average of all records made in that herd in the particular year the record was made. This method restricts the study to differences between near-

contemporary records made in the same herd. Differences between herd averages and between yearly averages in the same herd are simply set aside, bypassed and left unanalyzed. Since, in each year, in each herd, a sufficiently large number of records were made, the average production for that year in that herd could be taken as a measure of the environmental effect for the year, if we could assume that the genetic part of differences between herds and between year averages within the same herd is negligible.

In the genetic analysis, both the sire and the dam components were desired. The available records were classified by the sire and by the dam of the cow which made them. A few sires and most of the dams had just one daughter each. Estimating differences between sire's or dam's breeding values from production records of just one daughter each would be inadequate. Therefore, all records that were made by a cow whose sire or dam had just one daughter were discarded from this part of the analysis. The remaining records were analyzed genetically in two ways: first, with the original 305 day M.E. 2X butterfat production and, second, with production corrected for environmental influences. The original record would show what happens when important environmental influences are ignored in genetic analyses. With very few exceptions a sire was used in only one herd. Each dam and her daughters produced in the same herd. The sums of squares

from the twelve herds were pooled. The analysis was done in three different ways: first, as a hierarchal classification by sires, dams within sires, among full sisters, and among records; second a similar hierarchal classification using full sisters only; and third, as a cross classification by sires and dams.

### Hierarchal classification

The model assumed for the analysis is:

$$y_{ijkl} = u + s_i + d_{ij} + c_{ijk} + e_{ijkl}$$

where,

$y_{ijkl}$  is the  $l$ th record of the  $k$ th daughter of the  $j$ th dam of the  $i$ th sire

$u$  is the general mean

$s_i$  is an effect common to the daughters of the  $i$ th sire

$d_{ij}$  is an effect common to the daughters of the  $j$ th dam when mated to the  $i$ th sire

$c_{ijk}$  is an effect peculiar to an individual cow but varying between full sisters

$e_{ijkl}$  is the effect of differences between records of the same cow

For the analysis to be valid, the above effects must be additive and independently distributed. In the genetic interpretation of the analysis, random mating was assumed. This may not be absolutely true but inbreeding was generally avoided in these state herds, where several bulls were often in use in a herd at the same time. The ideals in mind when

choosing bulls and culling cows were fairly constant and any serious departure from random mating seems highly unlikely.

The model is similar to one described by Kempthorne (1955). He gives, in this paper, the following theoretical values for the different components:

V(S) - the variance due to genetic differences between sires - is equal to covariance between paternal half sisters which is an estimate of  $1/4 \sigma_A^2 + 1/16 \sigma_{AA}^2 + 1/64 \sigma_{AAA}^2 + \dots$ . Here, the first term is the variance due to additive gene effects and the other terms are the variance due to some of the epistatic effects. If the epistatic effects are negligible, then four times this component is the additive genetic variance in the population.

V(D) - the variance due to differences between dams within sires - is the covariance between full sisters minus the covariance between half sisters. It contains  $1/4 \sigma_A^2 + 1/4 \sigma_D^2 + 3/16 \sigma_{AA}^2 + 1/8 \sigma_{AD}^2 + 1/16 \sigma_{DD}^2 + \dots$ ; where the first term is the additive variance, the second term is the dominance variance and the other terms are parts of the epistatic variance. If epistatic effects are negligible and maternal effects are absent, then four times this component gives the total additive variance plus the dominance variance in the population.

The dam component minus the sire component is an estimate of one-fourth the dominance variance, also assuming negligible

epistatic variance. However, to the extent that environmental factors are confused with genetic differences or, epistatic differences are not negligible, to that extent these estimates may be exaggerated.

Similar techniques have been used to get components of genetic variance in corn and poultry. In the case of living organisms which produce several offspring within a short time and where those offspring can be kept in a common environment, one can get balanced observations in different subgroups and expectations of mean squares are easily obtained. Hazel and Lamoreux (1947) conducted such a planned experimental investigation in poultry on two characteristics: sexual maturity and body weight. They obtained estimates of "nicking" effects, maternal effects and heritability. Comstock and Robinson (1948) and Robinson et al. (1955) estimated additive and dominance effects from a similar hierarchal classification in corn.

The analysis of variance is given in Table 15. By equating the expected sums of squares to the observed sums of squares, using the coefficients shown in Table 16, and solving the resulting simultaneous equations, the components shown in Table 17 were obtained.

The total variance among the original records was 7787, while the corresponding figure with the corrected records was 6841. The difference is what was removed by taking each

Table 15. Analysis of variance: hierarchal classification

Source	d/f	Original records		Corrected records	
		S.S.	M.S.	S.S.	M.S.
Sires	283	7,645,081	27,014	5,062,443	17,888
Dams within sires	1,740	19,876,180	11,423	18,479,092	10,620
Full sisters	363	3,392,582	9,346	3,244,989	8,939
Records	4,605	23,028,735	5,001	20,787,738	4,514

Table 16. Coefficients in the expectations of sums of squares: hierarchal classification

Source	V(S)	V(D)	V(C)	V(E)
Sires	6362	1114	982	283
Dams within sires	-	5824	5056	1740
Full sisters	-	-	914	363
Records	-	-	-	4605

record as a deviation from that year's average within that herd. Most of this difference was in the component for differences between records of the same cow. This is to be expected since two records from the same cow are always made in different years and, therefore, this component in the original

Table 17. Components of variance: hierarchal classification

Source	<u>Original records</u>		<u>Corrected records</u>	
	Value	Per cent	Value	Percent
V(S)	639	8.2	271	4.0
V(D)	421	5.4	299	4.4
V(C)	1726	22.2	1757	25.7
V(E)	<u>5001</u>	<u>64.2</u>	<u>4514</u>	<u>66.0</u>
Total	7787	100.0	6841	100.0

records would always contain an intraherd year difference. However, both sire and dam components -  $V(S)$  and  $V(D)$  - are also smaller with the corrected records. Percentagewise they decrease even more than  $V(E)$  does. This suggests that environmental factors are not randomly distributed among sires and dams. Consequently environmental components are included in  $V(S)$  and  $V(D)$  in the analysis of the original records. This exaggerates the estimates of genetic components made from the original records. However, the process of expressing every record as a deviation of the herd-year average also removed from the corrected records the genetic differences which existed between herds and between years within herds. Presumably these were small.

The estimate of additive variance -  $4 V(S)$  - is 16.0 per cent while the estimate of additive plus dominance variance -  $4 V(D)$  - is 17.6 per cent, giving evidence that dominance variance is probably of minor importance in dairy cattle breeding. The estimate of heritability (narrow sense) is thus 16.0 per cent. Repeatability, which is an estimate of the total genetic variance plus the variance due to permanent environmental effects on the records of a cow, can be obtained as the ratio,  $\frac{V(S) + V(D) + V(C)}{V(S) + V(D) + V(C) + V(E)}$ , which in this investigation is 34.1 per cent.

Full sister analysis

A substantial number of full sister groups were available in the investigation, as can be surmised from the next-to-the-bottom line in Table 15. A similar analysis of variance was done on the records of these cows. This analysis is parallel to the preceding one but only includes the sets of full sisters. The covariance between full sisters is an estimate of  $1/2 \sigma_A^2 + 1/4 \sigma_D^2 + 1/4 \sigma_{AA}^2 + 1/16 \sigma_{DD}^2 + 1/8 \sigma_{AD}^2 + \dots$ , which reduces to  $1/2 \sigma_A^2 + 1/4 \sigma_D^2$ , if epistatic effects are negligible. Multiplying this quantity by two gives an estimate of the total additive variance plus half of the dominance variance in the population.

The analysis of variance is given in Table 18. The components were obtained using the coefficients in Table 19 and are shown in Table 20.

The reduction in total variance when the records are corrected is about the same in this analysis as in the previous one shown in Table 17. While most of this is in the error term, a substantial drop is in the component for differences between full sister groups. This suggests, as before, that environmental effects were correlated as between full sisters and, therefore, any genetic interpretation which assumes that in the original records  $V(\text{FS})$  is wholly genetic is certain to give exaggerated results.

Twice the full sister group component -  $2 V(\text{FS})$  - which

Table 18. Analysis of variance: full sisters

Source	d/f	Original records		Corrected records	
		S.S.	M.S.	S.S.	M.S.
Full sister group	306	5,052,067	16,510	4,019,683	13,136
Full sisters within groups	363	3,392,582	9,346	3,244,989	8,939
Records	<u>1,275</u>	<u>6,892,514</u>	5,406	<u>6,274,561</u>	4,921
	1,944	15,337,163		13,539,233	

Table 19. Coefficients of the components in the sums of squares: full sisters

Source	V(FS)	V(C)	V(E)
Full sister groups	1950	1039	306
Full sisters within groups	-	913	363
Records	-	-	1275

is 12.6 per cent in the corrected records, is an estimate of the total additive variance plus half of the dominance variance among contemporary cows in the same herd. While, the estimate should actually be larger than the estimate of additive variance only, obtained in the earlier analysis, the estimate here is less than the 16.0 per cent obtained in the previous section. Only about a fourth of the information

Table 20. Components of variance: full sisters

Source	<u>Original records</u>		<u>Corrected records</u>	
	Value	Per cent	Value	Per cent
V(FS)	908	11.5	438	6.3
V(C)	1567	19.9	1598	23.0
V(E)	<u>5406</u>	<u>68.6</u>	<u>4921</u>	<u>70.7</u>
Total	7881	100.0	6957	100.0

used in the former analysis is available in this analysis. Thus the discrepancy could easily be due to sampling error. That the two estimates are fairly close suggests again that the dominance variance is of minor importance in cattle breeding.

The component,  $V(C)$ , obtained in the above analysis is composed of  $\sigma_{EP}^2 + 1/2 \sigma_A^2 + 3/4 \sigma_D^2 + 3/4 \sigma_{AA}^2 + 15/16 \sigma_{DD}^2 + 7/8 \sigma_{AD}^2 + \dots$ , where the first term refers to permanent effects of individual cow environmental differences, such as the effects of mastitis, injuries to the udder, etc., the second term to additive variance, the third term to dominance variance and the other terms to epistatic variance. This component,  $V(C)$ , minus  $V(FS)$  is an estimate of  $\sigma_{EP}^2 + 1/2 \sigma_D^2 + 1/2 \sigma_{AA}^2 + 7/8 \sigma_{DD}^2 + 3/4 \sigma_{AD}^2 + \dots$ . An estimate of  $\sigma_D^2$  was obtained in the earlier analysis and using this figure and  $V(FS)$  and  $V(C)$  as shown in Table 20, a value of 15.9 per cent is obtained as an estimate of the combined effects of  $\sigma_{EP}^2$  and a large share of the epistatic variance.

The mean squares due to records in Table 18 should be expected to agree with that in Table 15. However, there is some difference, the records of full sisters having slightly higher variation than those of all cows, which includes these full sisters. This difference could have been due to chance. It is also possible, that the full sister sets may

to some extent, represent animals highly prized on account of their pedigrees. In that case, they might have been retained in the herds longer even though they had some disappointing small records.

### Cross classification

This analysis was of the same information used in the hierarchal classification. This model differs from the hierarchal one only in that the component obtained for dams within sires is divided here into a dam component and an interaction component due to sires and dams. If the factors in the model are correlated, then the estimates of main effects and interactions are biased. Under such circumstances the significance of interactions cannot be interpreted. If covariance between the effects of sires and dams exists, such covariance is credited to the sires and then credited again to the dams. This covariance is taken out twice from the interaction figure, which results in the interaction being too small if the covariance is positive and too large if it is negative. If there is correlation in this investigation, it is probably a small positive value if there was any regular trend toward genetic improvement over the period studied. If maternal effects are unimportant, and sire and dam are uncorrelated, the estimates of both sire and dam components have the same expected theoretical

Table 21. Components of variance: cross classification

Source	Original records		Corrected records	
	Value	Per cent	Value	Per cent
V(S)	643	8.2	273	4.0
V(D)	458	5.8	173	2.5
V(SD)	- 37	-	126	1.8
V(C)	1725	22.0	1757	25.7
V(E)	<u>5001</u>	<u>63.9</u>	<u>4514</u>	<u>66.0</u>
Total	7827		6843	

values of  $1/4 \sigma_A^2 + 1/16 \sigma_{AA}^2 + \dots$ . The sire by dam interaction is a measure of the deviation of the production of daughters of the particular sire and dam on the basis of their progeny averages. If the average production of several daughters of a sire, from a random sample of dams, is 40 pounds of fat above the herd average and if the average production of several daughters of a dam, from a random sample of sires, is 20 pounds of fat above herd average, then the expected average production of daughters of this particular dam by this sire would be 60 pounds above the herd average. The deviation from these expected results could be due to dominance, to epistatic effects, to environment and genotype interactions, and to sampling errors.

The components obtained from this analysis are given in Table 21. The estimate of sire by dam interaction is 1.8 per cent, a figure of low magnitude. This suggests that deviations from the additive scheme of gene effects are probably unimportant. However, the breeding values of dams and of a few sires are obtained on the performance of only a few daughters, mostly two. Consequently the amount of information available concerning the interaction component is too small for it to be very accurately estimated.

#### Confidence limits of components

The components obtained in this analysis are statistics from a sample. One desires to make some inferences about the parameters in the population which this sample represents. As indicated earlier, if the distribution of the variates were normal, in addition to the other assumptions made, confidence limits on parameters in the population can be obtained from these estimates.

Bross (1950) and Crump (1951) dealt with the principles and methods involved in getting fiducial intervals for various components. The variance of a mean square is twice the square of that mean square divided by its degrees of freedom plus two. If a mean square is designated  $V_1$ , then the variance of  $V_1$  is equal to  $\frac{2 V_1^2}{f_1 + 2}$ , where  $f_1$  is the degrees of freedom for  $V_1$ . An analysis of variance contains a few mean squares,

each a linear function of one or more components of variance. From the variances of linear functions, one can obtain the variances of these components. For instance, in the following analysis of variance:

Source	d/f	M.S.	Exp. M.S.
Between groups	a	A	$\sigma_e^2 + K \sigma_g^2$
Within groups, error	b	B	$\sigma_e^2$

$$\text{then, Var. (B)} = \frac{2 B^2}{b + 2}$$

$$\text{and since } \hat{\sigma}_e^2 = B,$$

$$\text{Var. (B)} = \text{Var. (} \hat{\sigma}_e^2 \text{)}$$

$$\text{now, } \hat{\sigma}_g^2 = \frac{A - B}{k} \text{ and}$$

$$\text{Var. (} \hat{\sigma}_g^2 \text{)} = \frac{1}{k^2} [\text{Var. (A)} + \text{Var. (B)}]$$

assuming that covariance between A and B is zero

$$= \frac{1}{k^2} \left[ \frac{2 A^2}{a + 2} + \frac{2 B^2}{b + 2} \right]$$

In a hierarchal classification the variance of each component can thus be obtained easily. The  $(1 - \alpha)$  confidence limits of the population parameter are the component obtained plus and minus  $T_\alpha$  times the standard error of the variance component. This standard error is the square root of the variance of the component. The confidence limits of the genetic components obtained in the analysis by hierarchal and full sister classifications are given in Tables 22 and 23.

Table 22. Confidence limits: hierarchal classification

Component	Value	Variance of component	Standard error	95 per cent confidence limits
V(S)	271	8,146	90.2	94 - 448
V(D)	299	65,432	255.8	0 - 800
V(C)	1757	70,344	265.2	1237 - 2277
V(E)	4514	8,846	94.1	4330 - 4698

Table 23. Confidence limits: full sister classification

Component	Value	Variance of component	Standard error	95 per cent confidence limits
V(FS)	438	49,890	223.4	0 - 876
V(C)	1598	74,924	273.7	1062 - 2134
V(E)	4921	37,927	194.7	4539 - 5303

Osborne and Paterson (1952) give a method of getting the variance of heritability estimates which are ratios of components of variance. For instance, if

$$z = \frac{A}{A + B + C},$$

then,

$$\text{Var. } z = \frac{A^2 \text{ Var. } (A+B+C) + (A+B+C)^2 \text{ Var. } A - 2A (A+B+C) \text{ Cov. } A, (A+B+C)}{(A + B + C)^4}$$

The standard errors for the heritability estimates were then obtained, assuming that the covariance between components (A, B, and C) is zero. The 95 per cent confidence limits of the heritability estimate obtained from the sire component are 6 to 26 per cent. Similarly, the 95 per cent limits on heritability obtained from the use of full sister records, which includes  $1/2 \sigma_D^2$ , are 1 to 24 per cent. The confidence limits of the heritability estimate should be considered with considerable caution, since there is no evidence for assuming that the variance components are independently and normally distributed.

## DISCUSSION OF RESULTS

The differences in milk production between individual cows are of general concern in commercial dairy husbandry, whether the differences are caused wholly by environment or by a combination of environment and genetic sources. The animal breeder, on the other hand, desires a measure of the genetic variation, since his breeding techniques, designed to make maximum genetic improvement in each generation by selection, differ under different circumstances.

The average yield of a herd in different periods does not give a dependable indication of the genetic advance achieved, because the effects of environment and of genetic time trends are confounded. Rendel and Robertson (1950) estimated that by the simplest and most direct way of selecting breeding stock purely on their phenotypes and mating at random, the expected rate of genetic improvement per year is 1.0 per cent of average milk production. Robertson and Rendel (1950) in another paper estimated that, with progeny testing and artificial insemination, a somewhat higher rate of genetic improvement could be attained, depending on heritability of the characteristic and size of the breeding unit, ranging from 1.26 to 2.06 per cent per year for heritability values ranging from 0.10 to 0.50 respectively.

With the widespread introduction of artificial

insemination, bulls are used more extensively and in many herds. While it is possible to progeny test bulls a little earlier in life by artificial insemination, this is at the cost of fewer bulls tested than under natural breeding. Thus, a greater degree of accuracy in estimating breeding value of a bull is needed to compensate for the less intense selection. Use of records made by daughters born in different herds introduces into the evaluation of the sire more environmental variations than will occur within herds. Some of the differences between records, after that due to inheritance is accounted for, are specific to a whole group of animals, such as those of the same herd, those made at the same year, etc. The remaining differences are random from cow to cow. If several daughters of a bull are considered, these random amounts tend to cancel or reduce each other but the sources specific to groups do not. These points are considered by Lush and McGilliard (1955) and by Carter (1956). The specific group effects could be reduced by providing standard environment or by correcting for these effects by proper statistical methods. The former is usually impractical either because of physical obstacles or because of the prohibitive cost involved. In Denmark, testing stations were designed to progeny test bulls by providing a standard environment in which daughters from different herds could perform, yet Robertson and Mason (1956) found extra

variation between progeny groups at the stations that was not repeated in the field. Generally the more practical method is to estimate these group effects and correct the records for those effects which are considered important but have not been sufficiently random over the sire groups being compared.

In this investigation, estimates were obtained on a few factors considered important and on their interactions. Herd differences accounted for about 6 per cent and herd by year interaction accounted for about 10 per cent of the total variance. All other sources, such as year effects, season effects and the other interactions, individually played a small role in the causes of variation and when put together accounted for only about 4 per cent of the total variance. These estimates on the important components disagree with many of the earlier studies. Plum (1934, 1935), with an investigation on Iowa D.F.I.A. fat production records on 2,316 cows from 95 herds spread over a period of 3 to 11 years, attributed 33 per cent of total variance to herd differences, 6 per cent to intraherd differences in feeding practices, and 3 per cent to differences in season of calving. Thus, these factors accounted for 42 per cent of the total variance in contrast to approximately 20 per cent obtained in this investigation for these factors and their interactions. Pirchner (1957) reported on genetic and environmental

differences between herds in butterfat production, using recent D.H.I.A. records from Iowa. In his analysis, with daughters obtained by artificial insemination, herd differences, year differences, and herd by year interaction accounted for 31, 3 and 8 per cent, respectively, of total variance. When he conducted the analysis within years, the herd differences, sire differences and herd by sire interaction accounted for 39, 6 and 3 per cent of total variance. His findings, therefore, agree fairly well with those of Plum.

Hickman and Henderson (1955) studied D.H.I.A. records of 3,912 cows located in 1,094 different herds in New York and by 126 different sires used artificially over a period of eight years. Herd differences in fat production accounted for 44.7 per cent of the total variance when studied within years. When the same study was repeated according to a general factorial arrangement of herds, sires and year-seasons, the components due to herds, year-seasons and herd by year-season interaction respectively were 32.8, 4.6 and 14.9 per cent. The year-season effect and herd by year-season interaction estimates are close to the estimates obtained in the present investigation. The herd component of 32.8 per cent is much higher than the 6 per cent obtained here, but is closer to the estimate of Plum (1935). Gaunt and Bartlett (1958) with 22,621 records made by 8,190 daughters

of 456 sires from 1952 to 1955 in 237 herds in New England and New York, reported that herd, season and year influences accounted for nearly half the variation in milk production. Wadell and McGilliard (1959) estimated that herd differences in three breeds, Jersey, Guernsey and Holstein-Friesian, accounted for 17, 33 and 21 per cent, respectively, of the total variance. Hofmeyr (1955), from an analysis of Danish data, concluded that variance between herds is responsible for about 25 per cent of the total variance.

In contrast to the above, Johansson and Hansson (1940) found 6.4 per cent of the total variance due to difference between herds. This estimate, obtained from production records of 3,000 cows from 13 herds of Swedish Red and White cattle over a period of 15 years, agrees closely with that obtained in the present investigation.

Herd differences can be due both to differences in feeding and management practices and to differences in the average genetic merit of the cows in the different herds. By a somewhat arbitrary, although plausible system of grading the scale of feeding in different herds, Plum (1935) estimated that 12 per cent of the total variance could be attributed to average differences between herds in feeding methods while the rest of the 33 per cent due to herd differences was caused by other environmental causes and by genetic differences. Pirchner (1957) and Pirchner and Lush

(1959) estimated that 10 per cent of differences between herds is genetic. In England, where commercial herds resort to purchase of bulls from a few elite herds, Robertson and Rendel (1954) pointed out that differences among herds in milk production account for 25 per cent of the total variance, all of which they attributed to environmental causes. McGilliard (1952) using data from Jersey Herd Improvement Registry herds which exchanged bulls, estimated that herd differences were composed of about one-third differences in average breeding value and about two-thirds differences in average management and environmental conditions.

The divergence of the herd component between the estimate from the present investigation and a few of the estimates from D.H.I.A. data referred above is very wide, which calls for some explanation. The present study related to twelve fairly large herds, all except one of which have been under a single administrative control, for most of the period, aimed at maintaining standard management and feeding conditions. This could have resulted in keeping the environmental differences between herds smaller than between herds in general. A single administration administering a common breeding policy would also tend toward making genetic differences between herds small. D.H.I.A. farm records used by other investigators, in contrast, included many small herds. With a greater number of individually owned small herds one is liable to come across greater management differences

between them. In some of these herds considerable attention must have been paid to having a few high producing cows, while in others average attention may have been paid. Some of these herds were specialized dairy farms while others were part of mixed farms.

A striking feature of this study was that the first lactation production averaged higher than other records. Possibly the records made in early life of a cow are over-corrected for age effects. It is also possible that many of the records made later in life are made under the effect of adverse conditions, such as mastitis or milk fever which may cause the culling of the cow. This could have played some part in the comparatively better performance of first records. The close agreement of the value of most components obtained by the two analyses, using all records and first records, suggests that these differences have not had serious effects on the analysis.

Another significant feature of the analysis is that the remainder mean squares is larger when all records are used than it is when only first records are used. Could this be mere sampling difference or could the difference be real? The confidence limits on these estimates were obtained for the classification by seasons. For all records the 95 per cent confidence limits are 6,306 and 6,636 while the similar confidence limits for first records are 5,729 to 6,317. The

overlap of these two confidence limits is so small that these two remainder mean squares seem likely to be really different. An F test on the ratio of the two mean squares gives the same conclusion. The first records are not affected by such factors as previous calving interval, length of previous dry period, or number of lactations completed. The difference between these mean squares may have resulted from the influence of the above factors in fat production.

Assumptions, such as additivity of effects and independence of variables, were made in the analysis to get components. Year and season effects are highly influenced by complex natural and economic phenomena such as climatic conditions, crop production, labor supply, etc. that one would have little reason to suspect correlations between these factors or between these factors and herd effects. However, one can expect some correlations among the variables in the genetic analysis. There is always some correlation between the cows mated to the same bull in any herd. Von Krosigk and Lush (1958) found that in one of these herds, herd no. 1, which was closed to outside breeding for 20 years there was also some inbreeding. The inbreeding among calves born in that herd in the same year averaged 7.4 per cent. Other herds in this investigation are of about the same size but were not closed to outside breeding and inbreeding was avoided. These herds, therefore, probably would have very

little inbreeding. With inbreeding and no dominance or epistasis, the additive variance within families is  $(1 - F)$  of the initial additive variance, where  $F$  is the coefficient of inbreeding. The sire component in herd no. 1, which had an average  $F$  of 0.07, therefore is expected to be  $\frac{1 - F}{4}$  of the additive variance in the initial population. Accordingly, in this herd the sire component should have been multiplied by  $\frac{4}{1 - F}$  (which = 4.3) instead of by 4 to estimate the additive variance in a non-inbred population. In the other herds where inbreeding was avoided, the problem could be correlation among dams mated to the same sire. In a one-sire herd with each sire unrelated to his predecessors and used only two years, as is often the case in small herds, the coefficient of relationship would seldom rise above 0.15. Freeman and Henderson (1959) found an average coefficient of additive relationship of 0.128 among related but non-inbred cows in eleven commercial herds in New York. This relationship among contemporary animals was 0.07. In the herds under investigation with many sires used at one time, this relationship would certainly be less. With a coefficient of relationship as high as 0.07 between dams, the covariance between half sisters will be  $(\frac{1 - 0.07}{4}) \sigma_A^2$ . To get the additive variance the multiplicative factor is 3.7 instead of 4. Thus if the sires' mates were related to each other, the

estimate of heritability obtained from the sire component is slightly higher than it should be. However, the difference would be so small that, in view of the fairly large sampling error involved in getting the estimate, the effect of this correlation between dams could be ignored.

In the genetic analysis, only the records of those cows whose sire or dam had at least two daughters were used. This might appear as introducing a bias in the analysis. However, this bias cannot be serious since this criterion eliminated both dams who were culled because of low production in their first lactation and those who remained in the herd for several lactations but had only one daughter in production.

Numerous heritability estimates have been obtained for butterfat production in dairy cattle. Earlier estimates were obtained by doubling dam-daughter correlations. Later, this was obtained by doubling intra-sire regression of daughter on dam to get rid of the effects of differences between herd means and between periods several years apart in the same herd. Johansson and Hansson (1940) estimated, by doubling daughter-dam correlations, that heritability for milk yield or for fat production was between 30 and 40 per cent. Lush and Straus (1942) estimated this statistic by doubling intra-sire regression using D.H.I.A. records in Iowa for the period 1936 to 1939. The average regression coefficient was 0.087

giving an estimate of heritability of 0.17. In that paper they showed how herd differences can inflate heritability figures. Lush et al. (1948) demonstrate how heritability can be obtained from variance components. Lush (1949) describes the principles involved in obtaining heritability in this manner. In this paper he deals with the problems of interpreting heritability. A heritability figure applies to a particular characteristic in a particular population. It is a ratio and consequently its value can change if either the numerator or denominator changes. All methods of estimating heritability involve measuring phenotypic likeness between related individuals. Any correlation between relatives due to environmental causes can bias estimates of heritability obtained.

Johansson (1954) obtained by variance component analysis heritability estimates of 0.56 for 250 day milk production from Danish data. He suspected this figure to be too high and speculated that differences between progeny groups in age at calving and the state of nutrition at the time the animals were brought to the testing stations probably were partly responsible for a high sire component. Rendel et al. (1957) obtained heritability estimates for milk production from English data by both daughter-dam regression and by variance component. By the former method, for six different breeds, the estimates ranged from 0.03 to 0.55 and by the

latter method from 0.09 to 0.43, most values ranging over 0.35.

Johansson (1953) gives evidence that the heritability estimates differ depending on the production level of the herds. For 12 low and 17 high producing herds in Swedish Red and White cattle his estimates of heritability of fat yield were 0.32 and 0.39, respectively. Mason and Robertson (1956) found similar results, heritability being 0.22 among high herds and 0.05 in the low herds, averaging 0.15. There is some question about whether these heritability estimates actually do differ according to the levels of production. These differences, in many cases, are so small that they could be due to sampling errors.

If there are real differences in heritability estimates, several explanations are possible. If the environmental part of the total variation in fat production differs under different circumstances, the denominator of the ratio estimating heritability varies and this could be responsible for differences in heritability estimates. In Johansson's (1953) paper the total mean squares for low herds and high herds, respectively, were 4,066 and 6,277. These figures were obtained with data, where each cow had five records. Both figures are considerably less than the total variance of over 8,000 found by Hickman and Henderson (1955). Possibly under European conditions the environmental differences in

milk production are less. Similarly the numerator, which is the covariance between relatives, possibly includes some covariance from environmental differences. To the extent that this is present, the heritability estimate will be inflated. In the present investigation, the heritability estimate obtained from the sire component was 0.16 when deviations from the herd average for that year were used. If the estimate had been made in the same manner with uncorrected records, the heritability estimate would have been 0.33. This is an example where the sire component, with the uncorrected records, contains a portion caused by covariance between environments.

This investigation suggests that deviations from the additive scheme of inheritance are unimportant. Seath and Lush (1940) looked for evidence of "nicking" in dairy cattle. "Nicking" is due to genes having in certain combinations effects very different from their average effects, i.e. to epistasis. In that case a sire proved on one group of females could have different value for his merit when proved on a genetically different group of females. All but one of the 13 sires studied by them had similar proof with groups of daughters by different sires. This led them to conclude (subject to the limitation that only a few sires were studied) that this study gave no indication that in proving sires "nicking" is important enough to need much

attention. Hazel and Lamoreux (1947) using components obtained by analysis of variance, found a small "nicking" effect, 1.2 and 1.4 per cent, for two characteristics, sexual maturity and body weight in poultry. Freeman and Henderson (1959) calculated from records of 1,210 Holstein animals in eleven commercial herds, the average additive and dominance relationships among related but non-inbred cows. The average additive relationship was 0.128 and the average dominance correlation was 0.007. From this, the authors concluded that the dominance portion of the total variance would contribute little to the total genetic covariance between relatives in commercial dairying, even if the total amount of dominance variance were large.

A striking feature of this and similar investigations is that much of the total variance remains unexplained. The residual mean square, after accounting for known environmental and genetic components, was 4,514 out of a total initial variance of approximately 8,200. That is, more than 50 per cent of the total variance in fat production remains unaccounted for. Part of this could be due to genotype and environment interactions. However, earlier work suggests that such interactions are small. Hickman and Henderson (1955) found that 2.4 per cent of total variance was due to herd by sire interaction. Wadell and McGilliard (1959) found negative values for the same and Legates et al. (1956)

got estimates of 0.0 to 1.2 for sire by herd interaction, studying three different breeds.

A major portion of variance left unaccounted for may be due to micro-environmental influences referred by McBride (1958). If these unknown factors are distributed at random over the factors considered in this investigation, the estimates of these factors remain unbiased; otherwise the estimates obtained in this investigation include parts of the effects of these unknown factors, depending on the kind and extent of the nonrandomness in the distribution of these unknown factors.

## SUMMARY

The data for this investigation came from twelve of fifteen state-owned herds in Iowa. The 12,623 records used were made by 4,487 cows during the period 1940 to 1956. The three smallest herds were not used in the investigation, since records were not available for these herds for certain years in the study.

The records were not distributed wholly normally and the variances among the twelve herds were not entirely homogeneous. However, since the purpose of this study was to estimate components of sources of variation rather than tests of significance, these findings do not limit the validity of the study seriously.

Components for three major environmental effects, herds, years, seasons and the interactions among them, herd by year, herd by season, and year by season, were estimated by Method 1 of Henderson (1953). Herd differences and herd by year interactions were responsible for about 6 and 10 per cent of the total variance, respectively. The other sources were of minor importance individually, together they accounted for 4 per cent of total variance. These estimates were highly consistent whether the study was made with all records or was made with first records only.

The records were corrected for the major environmental effects by taking each record as a deviation from the average

of all records made in that herd in the same year the record was made. The genetic analysis was then made, using only the records of cows whose sires and dams each had at least two daughters in the herd. This restricted the genetic analysis to 7,003 records by 2,398 daughters of 295 sires. This analysis was made in three different ways: (1) hierarchal classification, (2) full sister analysis, and (3) cross classification. By the first method the intra-herd heritability of butterfat production was estimated as 0.16, assuming epistatic effects to be negligible, being four times the sire component. By the second method the estimate, as twice the full sister component, was 0.12. By the first method 1.6 per cent of total variance was attributed to dominance deviations and other forces such as maternal effects, that make full sisters resemble more than paternal half-sisters. By the last method the component for sire by dam interaction was 1.8 per cent. The sire by dam interaction measures deviation from the average effects of genes and could include variance due to dominance, to epistatic effects, to environment and genotype interaction and to sampling errors. The small magnitude of this estimate and the 1.6 per cent of total variance as due to dominance deviations, suggest that deviations from an additive scheme of gene effects are probably not important, although they may be real. These estimates were obtained

under certain assumptions, the validity of which, often, cannot be substantiated.

When the genetic analysis was conducted with original records uncorrected for environmental effects (i.e. not expressed as deviations from the herd-year average), considerably higher estimates of genetic components were obtained. This happened because the herd and herd-year effects were not randomly distributed with respect to sires and dams. If, as seems probable, the genetic part in those differences between herds and between years within the same herd was very small, the resulting estimates of genetic components contained a considerable portion which was really environmental in origin, even though it appeared as a "sire component" or a "dam component." This might partially explain the divergence in heritability estimates obtained for butterfat production in various investigations.

An impressive feature of the investigation is that after the estimates of environmental and genetic components were made, a little more than 50 per cent of the total variance remains unexplained. Presumably most or all of this is caused by unknown environmental variations not easily defined. If these unknown factors were randomly distributed among the named factors studied, then the estimates obtained in this investigation are unbiased.

## LITERATURE CITED

- Anderson, R. L. and Bancroft, T. A. 1952. Statistical theory in research. New York, McGraw-Hill Book Co., Inc.
- Bross, I. 1950. Fiducial intervals for variance components. *Biometrics* 6:136-144.
- Cannon, C. Y. 1933. Seasonal effect on yield of dairy cows. *J. Dairy Sci.* 16:11-15.
- Carter, A. H. 1956. Some aspects of dairy sire selection. *Proc. N. Z. Soc. Anim. Prod.* 16:77-88.
- Cochran, W. G. 1947. Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics* 3:22-38.
- Comstock, R. E. and Robinson, H. F. 1948. The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. *Biometrics* 4:254-266.
- Copeland, L. 1934. Milk and butterfat yields of Jersey cows as affected by frequency of milking. *J. Dairy Sci.* 17:815-821.
- Crump, S. L. 1946. The estimation of variance components in analysis of variance. *Biometrics Bull.* 2:7-11.
- \_\_\_\_\_. 1951. The present status of variance component analysis. *Biometrics* 7:1-16.
- Darwin, C. 1875. The variation of animals and plants under domestication. 2nd ed. Vol. I. London, John Murray.
- East, E. M. and Jones, D. F. 1919. Inbreeding and outbreeding. Philadelphia, J. B. Lippincott Company.
- Eckles, C. H. 1915. The ration and age of calving as factors influencing the growth and dairy qualities of cows. *Missouri Agr. Expt. Sta. Bul.* 135.
- Eisenhart, C. 1947. The assumptions underlying the analysis of variance. *Biometrics* 3:1-21.

- Fisher, R. A. 1958. The genetical theory of natural selection. 2nd ed. New York, Dover Publications, Inc.
- Fohrman, M. H. 1926. Factors for adjusting milk and butterfat records of Register of Merit Jersey cows to a uniform age basis. *J. Dairy Sci.* 9:469-480.
- Freeman, A. E. and Henderson, C. R. 1959. Genetic structure of dairy cattle herds in terms of additive and dominance relationships. *J. Dairy Sci.* 42:621-625.
- Gaines, W. L. and Palfrey, J. R. 1931. Length of calving interval and average milk yield. *J. Dairy Sci.* 14:294-306.
- Gaunt, S. N. and Bartlett, M. A. G. 1958. Influence of herd and stablemate averages on production records of artificially sired (A. I.) progeny. *J. Dairy Sci.* 41:341.
- Gifford, W. and Crosby, J. E. 1933. The effect of frequency of milking on the annual production of dairy cows. *Missouri Agr. Expt. Sta. Bul.* 328:21.
- Gowen, J. W. 1934. The influence of inheritance and environment on the milk production and butterfat percentage of Jersey cattle. *J. Agr. Res.* 49:433-435.
- Hammond, J. and Sanders, H. G. 1923. Some factors affecting milk yield. *J. Agr. Sci.* 13:74-119.
- Hardy, G. H. 1908. Mendelian proportions in a mixed population. *Science* 28:49-50.
- Hazel, L. N. and Lamoreux, W. F. 1947. Heritability, maternal effects and nicking in relation to sexual maturity and body weight in White Leghorns. *Poultry Sci.* 26:508-514.
- Henderson, C. R. 1953. Estimation of variance and covariance components. *Biometrics* 9:226-252.
- Hickman, C. G. and Henderson, C. R. 1955. Components of the relationship between level of production and rate of maturity in dairy cattle. *J. Dairy Sci.* 38:883-890.

- Hofmeyr, J. 1955. A study of Danish and Swedish progeny testing methods for dairy bulls. Kungl. Lantbrukshogskolans Annaler 22:425-488.
- Jennings, H. S. 1916. The numerical results of diverse systems of breeding. Genetics 1:53-89.
- Johansson, I. 1953. The manifestation and heritability of quantitative characters in dairy cattle under different environment conditions. Acta Genet. 4:221-231.
- \_\_\_\_\_. 1954. An analysis of data from the Danish bull progeny testing stations. Zeitschrift fur Tierzuchtung und Zuchtungsbiologie 63:105-126.
- \_\_\_\_\_ and Hansson, A. 1940. Causes of variation in milk and butterfat yield of dairy cows. Kungl. Lantbruksakademiens Tidskrift 79(6 1/2):1-127.
- Kempthorne, O. 1955. The theoretical values of correlations between relatives in random mating populations. Genetics 40:153-167.
- \_\_\_\_\_. 1957. An introduction to genetic statistics. New York, John Wiley and Sons, Inc.
- Kendrick, J. F. 1955. Standardizing Dairy Herd Improvement Association records in proving sires. U.S. Agr. Research Series No. 52-1.
- Legates, J. E., Verlinden, F. J., and Kendrick, J. F. 1956. Sire by herd interaction in production traits in dairy cattle. J. Dairy Sci. 39:1055-1063.
- Lerner, M. 1958. The genetic basis of selection. New York, John Wiley and Sons, Inc.
- Lush, J. L. 1940. Intra-sire correlations or regression of offspring on dams as a method of estimating heritability of characteristics. Proc. Amer. Soc. Anim. Prod. 1940:293-301.
- \_\_\_\_\_. 1945. Animal breeding plans. 3rd ed. Ames, Iowa, Iowa State College Press.
- \_\_\_\_\_. 1949. Heritability of quantitative characters in farm animals. Hereditas (Suppl. Vol.) 1949:356-375.

- \_\_\_\_\_. 1958. Dairy cattle genetics. In Frandsen, J. H. ed. Dairy herdbook and dictionary. pp. 22-25. Amherst, Mass., J. H. Frandsen.
- \_\_\_\_\_, Lamoreux, W. F., and Hazel, L. R. 1948. The heritability of resistance to death in the fowl. Poultry Sci. 27:375-388.
- \_\_\_\_\_ and McGilliard, L. D. 1955. Proving dairy sires and dams. J. Dairy Sci. 38:163-180.
- \_\_\_\_\_ and Shrode, R. R. 1950. Changes in milk production with age and milking frequency. J. Dairy Sci. 33:338-357.
- \_\_\_\_\_ and Straus, F. S. 1942. The heritability of butterfat production in dairy cattle. J. Dairy Sci. 25:975-982.
- Mason, I. L. and Robertson, A. 1956. The progeny testing of dairy bulls at different levels of production. J. Agr. Sci. 47:367-375.
- McBride, G. 1958. The environment and animal breeding problems. Anim. Breeding Abs. 26:349-358.
- McGilliard, L. D. 1952. Usefulness of the herd average in estimating breeding values of dairy cattle. Unpublished Ph. D. Thesis. Ames, Iowa, Library, Iowa State University of Science and Technology.
- Osborne, R. and Paterson, W. S. B. 1952. On the sampling variances of heritability estimates derived from variance analyses. Royal Society of Edinburgh Proc. 64:456-461.
- Pearl, R. 1913. A contribution towards an analysis of the problem of inbreeding. Am. Naturalist 47:577-614.
- \_\_\_\_\_. 1914a. On the results of inbreeding a Mendelian population. A correction and extension of previous conclusions. Am. Naturalist 48:57-62.
- \_\_\_\_\_. 1914b. On a general formula for the constitution of the nth generation of a Mendelian population in which all matings are of brother x sister. Am. Naturalist 48:491-494.
- Pirchner, F. 1957. Genetic and environmental differences between herds in butterfat production. Unpublished

Ph. D. Thesis. Ames, Iowa, Library, Iowa State University of Science and Technology.

- \_\_\_\_\_ and Lush, J. L. 1959. Genetic and environmental portions of the variation among herds in butterfat production. *J. Dairy Sci.* 42:115-122.
- Plum, M. M. 1934. The relative importance of heredity and environment in determining the butterfat production of cows in Iowa Cow Testing Associations. Unpublished Ph. D. Thesis. Ames, Iowa, Library, Iowa State University of Science and Technology.
- \_\_\_\_\_. 1935. Causes of differences in butterfat production of cows in Iowa Cow Testing Associations. *J. Dairy Sci.* 18:811-825.
- Rendel, J. M. and Robertson, A. 1950. Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. *J. Genetics* 50:1-8.
- \_\_\_\_\_, \_\_\_\_\_, Asker, A. A., Khishin, S. S., and Ragab, M. T. 1957. The inheritance of milk production characteristics. *J. Agr. Sci.* 48:426-432.
- Robbins, R. B. 1917. Applications of mathematics to breeding problems I. *Genetics* 2:489-504.
- \_\_\_\_\_. 1918a. Application of mathematics to breeding problems II. *Genetics* 3:73-92.
- \_\_\_\_\_. 1918b. Some applications of mathematics to breeding problems III. *Genetics* 3:375-389.
- Robertson, A. and Mason, I. L. 1956. The progeny testing of dairy bulls - A comparison of special station and field results. *J. Agr. Sci.* 47:376-381.
- \_\_\_\_\_ and Rendel, J. M. 1950. The use of progeny testing with artificial insemination in dairy cattle. *J. Genetics* 50:21-31.
- \_\_\_\_\_ and \_\_\_\_\_. 1954. The performance of heifers got by artificial insemination. *J. Agr. Sci.* 44:184-192.
- Robinson, H. F., Comstock, R. E., and Harvey, P. H. 1955. Genetic variances in open pollinated varieties of corn. *Genetics* 40:45-60.

- Sanders, H. G. 1927. The variations in milk yields caused by season of the year, service, age and dry period and their eliminations. Part I. Season of the year and Part II. Service. J. Agr. Sci. 17:339-379 and 502-523.
- \_\_\_\_\_. 1928. The variations in milk yields caused by season of the year, service, age and dry period and their elimination. Part III. Age and Part IV. Dry period and standardization of yields. J. Agr. Sci. 18:46-67 and 209-251.
- Seath, D. M. and Lush, J. L. 1940. "Nicking" in dairy cattle. J. Dairy Sci. 23:103-113.
- Snedecor, G. W. 1956. Statistical methods. 5th ed. Ames, Iowa, Iowa State College Press.
- Turner, C. W., Ragsdale, A. C., and Brody, S. 1924. The relation between age, weight and fat production in dairy cows. Missouri Agr. Expt. Sta. Bul. 221.
- U. S. Agricultural Marketing Service. 1956. Milk - Farm production, disposition and income. Crop Reporting Board Statistics Bul. 175:4.
- Von Krosigk, C. M. and Lush, J. L. 1958. Effects of inbreeding on production in Holsteins. J. Dairy Sci. 41:105-113.
- Wadell, L. H. and McGilliard, L. D. 1959. Influence of artificial breeding on production in Michigan dairy herds. J. Dairy Sci. 42:1079-1085.
- Wentworth, E. N. and Remick, B. L. 1916. Some breeding properties of the generalized Mendelian populations. Genetics 1:608-616.
- Wylie, C. E. 1925. The effect of season on the milk and fat production of Jersey cows. J. Dairy Sci. 8:127-131.

## ACKNOWLEDGMENTS

The writer records, with great pleasure, his gratitude to Dr. J. L. Lush for the guidance and inspiration obtained from him during the two years of graduate study, and for the time spent in consultation, from time to time, on the progress of this investigation and writing this dissertation. Sincere appreciation for the time spent by Dr. A. E. Freeman in consultation on the investigation is expressed.

But for the Fellowship from the Rockefeller Foundation, leave of absence from the Allahabad Agricultural Institute, and permission to use and process data from state herds by the Department of Animal Husbandry of Iowa State University, this fruitful experience of the writer would not have been possible. For these opportunities the writer is exceedingly grateful.