V.80 V88 V/95

V

SAMPLING OF MILK RECORDS FOR ESTIMATING LACTATION YIELD

By

B.S. Gill

V/2K

Dissertation submitted in partial fulfillment of
the requirements for the award of Diploma
in Agricultural and Animal Husbandry
Statistics of the Indian Council
of Agricultural Research

NEW DELHI

1957

DE-04/A

CONTENTS

ACKNOWLEDGEMENTS

I have pleasure in expressing my deepest sense of gratitude to Shri V.N. Amble, Statistician Animal Husbandry, Indian Council of Agricultural Research for his valuable guidance, constant encouragement and constructive criticism during the course of investigation and writing this thesis.

My thanks are due to Dr. V.G. Panse, Statistical Adviser, ICAR, New Delhi, for his many helpful suggestions and for providing facilities for research work.

B.S. Gill

1. INTRODUCTION

One of the most important problems in Applied Statistics in the field of Animal Husbandry is the evolution of a technique for assessing the progress made under different development schemes for improving livestock particularly in respect of milk production.

The improvement in the productive capacity of cattle is being pursued through such schemes as Key Village Scheme and cross breeding schemes in hilly and heavy rainfall areas. Under a scheme of this nature the village cattle in a compact area are sought to be intensively developed through improved methods of breeding, feeding and rearing.

For the assessment of the progress in the augmentation of milk production in any such scheme it is necessary to estimate objectively the productive capacity of the animals at given periods and of those belonging to different generations. It is desirable to base this estimation on the estimates of lactation yields of a representative set of animals rather than on a sample of all the daily records made on given days (taking independent samples on different days), since the lactation yields provide a standard basis for making comparisons between groups of animals as are required for studying generation to generation differences, the breeding worth of sire etc.

Maintenance of daily milk yield records of all cows in an area for the purpose of getting the lactation yields is not practicable because the responsibility of such recording cannot be put on the owners themselves since they are often not literate and are not yet in a position to appreciate the value of accurate recording. It would be necessary in the present circumstances to collect the information by employing trained investigators to record the milk yields. In such a case complete daily recording of all the cows is prohibitive in cost and recourse has to be taken to sampling of cows as well as days in milk of selected cows. It is, therefore, necessary to evolve a suitable procedure of sampling of a group of cows and of their daily milk yields, and of working out an efficient estimate of the average lactation yield per animal of the group.

In the absence of any available data on daily yields of cows from rural areas, it was considered worthwhile to make a start in the study of this type of problem with the data on daily milk yields of cows of Hariana breed from Government Livestock Farm, Hissar, Punjab, collected in a scheme of the Indian Council of Agricultural Research.

The investigation of the problem was initiated by Shri Khandekar, a diploma student of the Council last year. He mainly concentrated on the sampling of daily milk yields for estimating the total lactation yield of a single cow. He studied the relation between the rate of milk secretion and the advance in lactation by fitting mathematical curves with a view to utilizing this information in increasing the efficiency of estimation.

This thesis deals with the further investigations carried out on the same problem during my assignment as a Senior Research Fellow during the year 1956-57. In this an attempt has been made to find the gain in efficiency of different methods of estimating the lactation yield of a single cow (with and without utilising the knowledge regarding the nature of lactation curve) from systematic samples. The number of cows required for estimating the average lactation yield, per cow, of a group of cows at a desired level of precision has been also worked out.

2. REVIEW OF LITERATURE

2.1. The attempt to record milk yields in an organized manner was first made in 1895 in Denmark by Emil Konradi who founded the Vejen Milk Recording Society.

Numerous workers have compared, for suitability and reliability, different lengths of interval of recording milk yields for estimating the total lactation yield of a cow. Two early workers-Flieshmann (1891) and Martiny (1895), (1912) and (1914)- compared monthly with weekly records. They found maximum errors of 22 and 12 per cent respectively and, together with a number of early European authors, concluded that monthly recording cannot give reliable results. Further work by American authors- Yapp (1919), Peterson (1925) and Copeland (1928)-indicated that this conclusion was premature.

M'Candlish and McVicar (1925) showed that the records should be regarded as being in the middle of the test period. A method based on this later came to be known as the 'centering date' method. The centering date is an arbitrary date set for each farmer at about the time of month that recorder will visit his herd. Production is then calculated by counting 15 days back from the centering date and 13-16 days ahead (including the centering date) according to the length of the month.

A large number of suggestions for recordings at non-systematic intervals or at particular points of the lactation have been proposed from time to time. Gaines(1927) and Cannon etal (1942) found that there was a high correlation between results from recordings made during fifth month and the total estimated from monthly recordings Saiz (1927) has proposed recording in the sixth week, fifth month and eighth month of the lactation.

The possibility of extending interval of recording daily milk yields to two months was first investigated in United States of America where large distances which have to be covered make recording at short intervals very expensive. McDowell (1927) found that average variation from actual production was 3.80 per cent for bimonthly as compared to 2.91 per cent for monthly interval of milk recording.

Houston et al (1932) found only in 5 per cent of cases the error in milk yield, calculated from recording at weekly intervals, from actual yield exceeded 2.5 per cent of the actual yield. They further found that maximum error in estimated milk yield from 14-day interval was about twice that for the weekly recordings.

Tylor and Chapman (1944) proposed the simple addition of first ten recording day values multiplied by 30.5 to give a figure for the 305-day lactation.

Daily recordings of 400 lactations were recalculated by Jordao et al (1947) to estimate yields from recording at various intervals from one to eight weeks; the maximum error for weekly interval of recordings fell within the limits -5 to +6 and that for biweekly recording between ±7 per cent - little greater than that found by them for weekly weighings. For 21-day interval of recording the maximum errors were -11 and +7 per cent while monthly and bi-monthly recordings gave errors of -11 to +12 percent and -14 to +12 per cent respectively.

In agreement with a number of other workers Dick(1950) observed a linear relationship between the length of interval and the error. The standard deviations of the percentage error rose from 0.37 for a 2-day interval to 2.39 for 28-day interval.

A comparison of various intervals from 30 to 150 days was carried out by Erb et al (1952). Errors of 2.5 and 3.4 were exceeded in 25 per cent of the records calculated from recordings at 30-day and 60-day intervals respectively. The corresponding level of errors for intervals of 90, 120 and 150 days were 5.6, 7.4 and 8.8 respectively.

Jardim et al (1956) dealt with the estimation of milk production by means of biweekly, monthly and bimonthly observations without taking into account the date of calving. The data studied were 72 records of cows of Holstien-Friesian breeds. They calculated the 'mean of deviations', standard error of mean for biweekly monthly and bimonthly weighings and concluded that biweekly and monthly weighings might be taken unbiased whereas the bimonthly recording overestimated the production by about 5 percent.

2.2. Khandokar (1956) initiated a detailed investigation into essentially statistical problems underlying the method of sampling and procedure of estimating the lactation yield. On the basis of lactation length he grouped the 41 Hariana cows of Hissar Livestock Farm, the daily milk yields, in first lactation of which were studied, into four groups. The first group consisted of cows having a lactation length less than 270 days. The second and the third groups comprised of cows with lactation lengths lying between 271 days to 300 days and between 301 days to 330 days respectively. The cows whose lactation lengths were greater than 331 days were put into the fourth group. The four groups consisted of 5, 13, 15 and 8 cows respectively. Out of the different types of curves fitted to describe the trend of daily milk yield with advance in lactation mainly two curves were found to be satisfactory. One was the polynomial of fourth degree which accounted for 85 per cent of the total variation for cows having lactation lengths less than 330 days. For cows of greater lactation lengths the same percentage of variation was accounted for by the fifth degree curve. The other curve was a quadratic-cum-log curve of the form

$$y = a_1 + a_2x + a_3x^2 + a_4\log_e x$$

More than 80 per cent of the variation in daily yields was accounted for by this curve in case of all the groups.

3. MATERIAL FOR STUDY

At the Livestock Farm Hissar Panjab some 200 and odd Hariana cows are under experiment in Livestock Research Investigation Scheme of the Indian Council of Agricultural Research, which is in progress there since 1944.

The daily morning and evening records of the milk yield of 64 cows in the first lactation constituted the material for the study of the present investigation.

Of the 64 first lactation records only 41 were amenable for precise statistical analysis, others being omitted on the grounds of incompleteness, illness etc.

At the farm the calves were not weaned at birth. Once in a week, however, the cow was stripped completely without the calf being allowed to suckle. In what follows these yields are known as 'stripped yields'. The trend shown by the stripped yield was nearly the same as that shown by daily yields, the only difference being that the level of production for stripped days was about 2 pounds higher than that for ordinary days. It was, therefore, considered permissible to replace the milk yields recorded on the day when the cow was completely stripped by the average of yields of three days preceding and three days succeeding the stripped day so as to simplify the theoretical and empirical investigation by bringing these yields on the same footing as the other records.

The daily milk yields were weighed correct to half a pound of milk.

For each cow the recording of daily milk yield was not made in the colostrum period which varied from 2 days to 9 days with an average of about 5 days. In the calculation of the lactation period and lactation yield this period was excluded.

The average lactation period of these 41 cows was 301 days. The average lactation yield and the average daily yield were 1201 and 3.9 pounds respectively. The coefficient of variation between daily yields within cows was 25 per cent.

4. SAMPLING OF MILK RECORDS FOR ESTIMATION OF LACTATION YIELD OF A SINGLE COW.

4.1. To begin with different procedures of sampling daily milk yield records and of estimating total lactation yield of a single cow were studied.

4.2. A study of the literature on the subject has revealed that numerous workers on similar problems have employed systematic sampling i.e. a method of recording milk yields at a constant interval starting with a randomly selected day falling within the first interval, for estimating the total production. There is no doubt that systematic sampling has a considerable practical advantage of being simple to execute. There are, however, certain objections to systematic sampling. One is that very inaccurate results will be obtained if there are any periodicity in the parent sequence and sampling interval is a multiple of basic period. In the present case, however, there is no periodicity in the relationship of daily milk yield and advance in lactation and, therefore, there is no risk involved in taking to systematic sampling. There is another disadvantage namely that there is no method corresponding to that which is available for different types of random sampling by which a valid estimate of the sampling error can be obtained from sample data themselves. On the other hand, if the efficiency of systematic sampling can be shown to be high as compared with random sampling from a study of complete lactation records it would be appropriate to recommend systematic sampling on a suitable scale adequate for estimating lactation yield with a desired level of precision.

4.3. Let N be the number of days in milk of a cow. Let y_{ij} denote the milk yield of a cow on ($i+\overline{j-1}\,k$)th day and x_{ij} the ($i+\overline{j-1}\,k$) th day itself where k is the interval of recording for a systematic sample, where $i \leq k$ is the random start. Further let $N = nk+r$ where $0 \leq r \leq k$. We can arrange all the N recordings in a two way table having k columns, the first r number of columns having $(n+1)$ recordings and the rest n .

The diagram shows the arrangement.

Columns

1	2	....	r	....	i	....	k
y_{11}	y_{21}	....	y_{r1}	....	y_{i1}	....	y_{k1}
y_{12}	y_{22}	....	y_{r2}	....	y_{i2}	....	y_{k2}
....	....	....	....	....	....	....	....
....	....	....	....	....	....	....	....
....	....	....	....	....	....	....	....
y_{1j}	y_{2j}	....	y_{rj}	....	y_{ij}	....	y_{kj}
....	....	....	....	....	....	....	....
....	....	....	....	....	....	....	....
y_{1n}	y_{2n}	....	y_{rn}	....	y_{in}	....	y_{kn}
$y_{1(n+1)}$	$y_{2(n+1)}$		$y_{r(n+1)}$				

We define $\bar{y}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$ where $n_i = n+1$ or n according as $i \leq$ or $> \gamma$,

and $t_i = \sum_{j=1}^{n_i} y_{ij}$ = Total of the ith column;

and $T = \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}$ = Total lactation yield.

Now the method of sampling consists of drawing a random number, say i, less than or equal to the interval of recording i.e. k and taking all recordings in the i-th column as recordings in the sample. Thus the systematic sample consists of milk yields $y_{i1}, y_{i2} \ldots\ldots y_{ij} \ldots\ldots y_{in_i}$.

Two different systems of estimation were tried, one without making use of the knowledge regarding the nature of the lactation curve and other with utilising this knowledge as secured by the studies earlier made by Khandokar.

4.4. First system of Estimation

Under this systems two methods were considered. One was to obtain the estimate of the total lactation yield by multiplying the total milk yield obtained from the systematic sample by the interval of recording i.e. k. Thus the estimate is

$$T_u = Kt_i = k \sum_{j=1}^{n_i} y_{ij}$$

The second method was to obtain the estimation of total lactation yield of a cow by multiplying the average daily milk yield calculated from the sample, by total number of days the cow was in milk i.e. N. The estimate

$$T_b = N\bar{y}_{i.}$$

Some of the properties of these two estimators are given in the next section.

4.5. Expected values of T_u and T_b.

a) The estimate $T_u = kt_i = k\sum_{j=1}^{n_i} y_{ij}$

$$E(T_u) = \frac{1}{k}\sum_{i=1}^{k} k\sum_{j=1}^{n_i} y_{ij} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij} = T$$

= Total lactation yield

Hence T_u is an unbiased estimate of the total lactation yield.

b) The estimate $T_b = N\bar{y}_{i.} = \frac{N}{n_i}\sum_{j=1}^{n_i} y_{ij}$

$$E(T_b) = \frac{1}{k}\sum_{i=1}^{k}\frac{N}{n_i}\sum_{j=1}^{n_i} y_{ij} \neq T \text{ necessarily}$$

Hence T_b is, in general, a biased estimate of the total lactation yield.

By definition bias $= E(T_b) - T$

$$= \frac{1}{k}\sum_{i=1}^{k}\frac{N}{n_i}\sum_{j=1}^{n_i} y_{ij} - \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}$$

$$= \frac{N}{k}\sum_{i=1}^{k}\bar{y}_{i.} - \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}$$

$$= N\sum_{i=1}^{k}\bar{y}_{i.}\left(\frac{1}{k} - \frac{n_i}{N}\right)$$

The bias will vanish either when $n_i = \frac{N}{k}$ i.e. when the lactation period is an exact multiple of the interval of recording, or when $\bar{y}_{i.}$ remains constant.

4.6. Variances of T_u & T_b

a) By definition $V(T_u) = E\left\{T_u - E(T_u)\right\}^2$

$$= E(T_u^2) - E^2(T_u)$$

Now $E(T_u^2) = E(k^2 t_i^2) = \frac{1}{k}\sum_{i=1}^{k} k^2 t_i^2$

$$= k\sum_{i=1}^{k} t_i^2 .$$

and $E^2(T_u) = T^2$

Therefore $V(T_u) = k\sum_{i=1}^{k} t_i^2 - T^2$

b) The estimate $T_b = N\bar{y}_{i.}$

$$E(T_b^2) = E(N^2\bar{y}_{i.}^2) = \frac{N^2}{k}\sum_{i=1}^{k}\bar{y}_{i.}^2$$

and $E^2(T_b) = \frac{N^2}{k^2}\left(\sum_{i=1}^{k}\bar{y}_{i.}\right)^2$

Therefore $V(T_b) = \frac{N^2}{k}\sum_{i=1}^{k}\bar{y}_{i.}^2 - \frac{N^2}{k^2}\left(\sum_{i=1}^{k}\bar{y}_{i.}\right)^2$

$$= \frac{N^2}{k}\left\{\sum_{i=1}^{k}\bar{y}_{i.}^2 - \frac{1}{k}\left(\sum_{i=1}^{k}\bar{y}_{i.}\right)^2\right\}$$

$$= \frac{N^2}{k}\left\{\sum_{i=1}^{k}(\bar{y}_{i.} - \bar{\bar{y}})^2\right\}$$

where $\bar{\bar{y}}$ is mean of the k column means .

4.7. Second system of Estimation

This system of estimation consists in fitting a curve of a known form to systematically recorded daily milk yields and summing the fitted function over the lactation length to estimate the total lactation yield of a cow.

For the reasons given in section 2.2 only two curves for this purpose were considered , one was the quadratic-cum-log curve of the form

$$y = a_0 + a_1x + a_2x^2 + a_3 \log_e x$$

where y is the milk yield, x the corresponding period of time in days excluding the days of colostrum, and a's are constants . The other curve examined was the polynomial of fourth degree of the form

$$y = b_0 + b_1x + b_2x^2 + b_3x^3 + b_4x^4$$

It will be appropriate to first give an outline of the method of fitting these curves and of obtaining the estimate of the lactation yield therefrom.

4.8. Quadratic-cum-log curve

Fitting of the curve :- Denoting x, x^2 and $\log_e x$ by three variables x_1, x_2 and x_3 respectively and measuring y and x's from their respective sample means ($\bar{y}$, $\bar{x}_1$, $\bar{x}_2$ and $\bar{x}_3$) the equation can be written in the form

$$Y = l_1 X_1 + l_2 X_2 + l_3 X_3$$

$$\text{where } Y = (y - \bar{y}) \text{ and } X_i = (x_i - \bar{x}_i)$$
$$\text{for } i = 1, 2 \ \& \ 3$$

It is in this form that the method of fitting the curve has been carried out by the method of least squares. This involves setting up of normal equations, inversion of a 3 X 3 matrix and evaluation of the constants (Fisher 1950).

Estimation of lactation yield

Suppose the fitted curve is

$$Y = l_1 X_1 + l_2 X_2 + l_3 X_3$$

$$\text{i.e. } y = \bar{y}_1 + l_1 (x_1 - \bar{x}_1) + l_2 (x_2 - \bar{x}_2) + l_3 (x_3 - \bar{x}_3).$$

Denoting by T_q the estimate of the total lactation yield of a cow obtained by summing the estimated daily milk yields over the whole lactation length, we have

$$T_q = N\bar{y}_1 + l_1 \sum_{x=1}^{N} (x_1 - \bar{x}_1) + l_2 \sum_{x=1}^{N} (x_2 - \bar{x}_2) + l_3 \sum_{x=1}^{N} (x_3 - \bar{x}_3) .$$

4.9. Fitting of the Polynomial Curve:- A similar procedure for fitting the polynomial curves could be adopted as for the curve described above, but in doing so the number of digits of certain sum of squares and sum of products in the normal equations which had to be retained gets too large beyond the capacity of the calculating machine. Therefore, this renders the inversion of matrix of rank equal to or higher than four extremely laborious and approximate. An alternative method of fitting polynomial curves with the help of orthogonal polynomials was, therefore, utilised. This method takes considerably less time in numerical calculations.

The use of orthogonal polynomial values in curve fitting is described in a number of places (Kendall, M.G) and is now so well known that only a brief sketch of the theory is given below.

Let us assume that the relation of daily milk yield η on the advance in lactation, say x days, is of the form

$$\eta = a_0 + a_1 x + \cdots\cdots\cdots + a_p x^p \quad \ldots \quad (1)$$

$$= f(x) \text{ say.}$$

The polynomial f(x) may be written in terms of a set of polynomials

$$\xi_0(x), \quad \xi_1(x) \quad \cdots\cdots\cdots \quad \xi_r(x) \quad \cdots\cdots \quad \xi_p(x)$$

where $\xi_r(x)$ is, for the present, an arbitrary chosen polynomial of degree r.

$$\eta = \phi(\xi) = \beta_0\xi_0 + \beta_1\xi_1 + \dots\dots\dots + \beta_p\xi_p \qquad (2)$$

Let y_{i1}, y_{i2} .... y_{in} ($n > p$) be the observations made on the values of η on the i-th, (i + k)th ....$(i + \overline{n-1}\,k)$th days of lactation. Then one can envisage fitting equation (1) or equivalent equation (2) to the observations to obtain an estimate of dependence of η on x .

Adopting the principle of least squares to estimate the parameters of the equation (2), and choosing the ξ polynomials for any set of values so that $\sum_{l \neq m}^{n} \xi_l \xi_m = 0$ where $\xi_0 = 1$ (3)

To get the estimate of $\beta_j = \dfrac{\sum^{n} \xi_j y}{\sum_{1}^{n} \xi_j^2} = B_j$ (say)

the summation extending over the sample values.

Since x's are equally spaced, they may be replaced by 0, 1, 2 ......(n-1) by suitable choice of origin and scale. Then ξ values are easy to calculate since they become functions of n only.

Let Z be a variate getting values 0, 1, 2, ...(n-1) then taking $\xi_0 = 1$ as defined above

and $\xi_1 = Z - \frac{n-1}{2}$

and determining the other ξ's by the recurrence relation

$$\xi_{r+1} = \xi_1\xi_r - \frac{r^2(n^2 - r^2)}{4(4r^2 - 1)}\xi_{r-1}$$

it can be shown that these ξ's satisfy relation (3) .

Ultimately all ξ 's become the functions of ξ_1 only. In particular

$$\xi_2 = \xi_1^2 - \frac{n^2-1}{12}$$

$$\xi_3 = \xi_1^3 - \frac{3n^2-7}{20}\,\xi_1$$

$$\xi_4 = \xi_1^4 - \frac{3n^2-13}{14}\,\xi_1^2 + \frac{3(n^2-1)\,(n^2-9)}{560}$$

$$\xi_5 = \xi_1^5 - \frac{5(n^2-7)}{18}\,\xi_1^3 + \frac{15n^4-230n^2+407}{1008}\,\xi_1$$

In order to convert these ξ values to integers in their lowest terms they are multiplied by a constant λ. The values corresponding to $\xi' = \lambda\,\xi$ are given by Fisher and Yates (1953) for fitting curves upto the fifth degree to observation ranging in number from 3 to 75. λ values are also given at the bottom of ξ' values.

b. <u>Estimation of lactation yield</u>

Let the days on which the milk yields are recorded be $1,\ 1+k,\ 1+2k, \ldots\ldots\ldots\ 1+\overline{n-1}\,k$. Then the variate $Z = \frac{x-1}{k}$ takes the values $0,\ 1,\ 2, \ldots\ldots\ldots (n-1)$.

As described above we can fit a polynomial

$$y = \bar{y}_{1.} + B_1\xi_1 + B_2\xi_2 + B_3\xi_3 + B_4\xi_4$$

By definition the estimate of total lactation yield is the sum of the estimated daily yields of milk over the entire range of x which varies from 1 to N.

Thus

$$T_p = N\bar{y}_{1.} + B_1\sum_{x=1}^{N}\xi_1 + B_2\sum_{x=1}^{N}\xi_2 + B_3\sum_{x=1}^{N}\xi_3 + B_4\sum_{x=1}^{N}\xi_4$$

The numerical calculation of $\sum_{x=1}^{N}\xi_j$ involves $\sum_{x=1}^{N}(\ \xi_1\ )^{j}$ for all values of $j = 0,\ 1,\ 2,\ \ldots\ldots\ j$.

By definition $\xi_1 = Z - \frac{n-1}{2} = \frac{1}{k}\left(x-1-\frac{n-1}{2}k\right)$

$$= \frac{1}{k}(x-d) \quad \text{where } d = i + \frac{n-1}{2}k$$

Therefore $\sum_{x=1}^{N}\xi_1 = \frac{1}{k}\sum_{x=1}^{N}(x-d) = \frac{1}{k}\left[\frac{N(N+1)}{2} - Nd\right]$,

$$\sum_{x=1}^{N}\xi_2 = \sum_{x=1}^{N}\xi_1^{2} - \frac{n^2-1}{12}N$$

$$= \frac{1}{k^2}\left[\frac{N(N+1)(2N+1)}{6} - dN(N+1) + Nd^2\right] - \frac{N\ (n^2 - 1}{12}$$

$$\sum_{x=1}^{N}\xi_3 = \sum_{x=1}^{N}\xi_1^{3} - \frac{3n^2-7}{20}\sum_{x=1}^{N}\xi_1$$

$$= \frac{1}{k^3}\left[\frac{N^2(N+1)^2}{4} - \frac{dN(N+1)(2N+1)}{2} + 3d^2\frac{N(N+1)}{2} - d^3N\right] - \frac{3n^2-7}{20k}\left[\frac{N(N+1)}{2} - \right.$$

and $$\sum_{x=1}^{N}\xi_4 = \sum_{x=1}^{N}\xi_1^{4} - \frac{3n^2-13}{14}\sum_{x=1}^{N}\xi_1^{2} + \frac{3(n^2-1)(n^2-9)}{560}N$$

$$= \frac{1}{k^4}\left[\frac{N(N+1)(2N+1)\{3N(N+1)-1\}}{30} - d\,N^2(N+1)^2 + d^2\,N(N+1)(2N+1) - 2d^3N(N+1)+1\right.$$

$$- \frac{3n^2-13}{14k^2}\left\{\frac{N(N+1)(2N+1)}{6} - d\,N(N+1) + Nd^2\right\} + \frac{3(n^2-1)(n^2-9)}{560}N$$

Thus we obtain T_p an estimate of total lactation yield.

4.10. Expected value of T_p

The estimate

$$T_p = N\bar{y}_{1.} + B_1 \sum_{x=1}^{N} \xi_1 + B_2 \sum_{x=1}^{N} \xi_2 + B_3 \sum_{x=1}^{N} \xi_3 + B_4 \sum_{x=1}^{N} \xi_4$$

we assume $N = nk$ for theoretical convenience

Taking the expectations, we get

$$E(T_p) = N\bar{y}_{..} + \left\{ E(B_1 \sum_{x=1}^{N} \xi_1) + E(B_2 \sum_{x=1}^{N} \xi_2) + E(B_3 \sum_{x=1}^{N} \xi_3) + E(B_4 \sum_{x=1}^{N} \xi_4) \right\}$$

Since $N\bar{y}_{..}$ is the actual total lactation yield, T_p is a biased estimate. The term in the curly brackets is the bias. We can obtain an upper limit to it in the following manner.

We define

$$\phi_0 = 1$$

$$\phi_1 = \frac{1}{k} \left(x - \frac{n+1}{2} k \right)$$

and ϕ_{r+1} by the recurrence relation

$$\phi_{r+1} = \phi_1 \phi_r - \frac{r^2 (n^2 - r^2)}{4(4r^2 - 1)} \phi_{r-1}$$

It can be easily shown that $\sum_{x=1}^{N} \phi_1 > \left| \sum_{x=1}^{N} \xi_1 \right|$

and in general $\sum_{x=1}^{N} \phi_r > \left| \sum_{x=1}^{N} \xi_r \right|$

Thus substituting $\sum_{x=1}^{N} \phi_r$ for $\sum \xi_r$, we get the upper limit of the bias

$$\text{upp.lim.Bias} = \sum_{x=1}^{N} \phi_1 \left| E(B_1) \right| + \sum_{x=1}^{N} \phi_2 \left| E(B_2) \right| + \sum_{x=1}^{N} \phi_3 \left| E(B_3) \right| + \sum_{x=1}^{N} \phi_4 \left| E(B_4) \right|$$

(since $\sum \phi_r$ is independent of the random start i.e. i).

$$= \sum_{x=1}^{N}\phi_1 \frac{|E(\sum^{n} y_{ij}\cdot\xi_1)|}{\sum^{n}\xi_1^2} + \sum_{x=1}^{N}\phi_2 \frac{|E(\sum y_{ij}\xi_2)|}{\sum^{n}\xi_2^2} + \sum_{x=1}^{N}\phi_3 \frac{|E(\sum y_{ij}\cdot\xi_3)|}{\sum^{n}\xi_3^2} + \sum_{x=1}^{N}\phi_4 \frac{|E(\sum y_{ij}\cdot\xi_4)|}{\sum^{n}\xi_4^2}$$

$$= \frac{\sum_{x=1}^{N}\phi_1}{\sum^{n}\xi_1^2}\left|\frac{1}{k}\sum_{i=1}^{k}\sum^{n} y_{ij}\,\xi_1\right| + \frac{\sum_{x=1}^{N}\phi_2}{\sum^{n}\xi_2^2}\left|\frac{1}{k}\sum_{i=1}^{k}\sum^{n} y_{ij}\cdot\xi_2\right| + \frac{\sum_{x=1}^{N}\phi_3}{\sum^{n}\xi_3^2}\left|\frac{1}{k}\sum_{i=1}^{k}\sum^{n} y_{ij}\cdot\xi_3\right| + \frac{\sum\phi_4}{\sum^{n}\xi_4^2}\left|\frac{1}{k}\sum_{i=1}^{k}\sum^{n} y_{ij}\xi_4\right|$$

$$= \frac{\sum_{x=1}^{N}\phi_1}{\sum^{n}\xi_1^2}\left|\sum^{n}\bar{y}_{\cdot j}\xi_1\right| + \frac{\sum_{x=1}^{N}\phi_2}{\sum^{n}\xi_2^2}\left|\sum^{n}\bar{y}_{\cdot j}\xi_2\right| + \frac{\sum_{x=1}^{N}\phi_3}{\sum_{1}^{n}\xi_3^2}\left|\sum\bar{y}_{\cdot j}\,\xi_3\right| + \frac{\sum_{x=1}^{N}\phi_4}{\sum^{n}\xi_4^2}\left|\sum\bar{y}_{\cdot j}\,\xi_4\right|$$

We would first get the order of $\sum_{x=1}^{N}\phi_r$ and of $\sum_{x=1}^{n}\xi_r^2$

($r = 1, 2, 3$ and 4)

It is well known that

$$\sum_{1}^{n}\xi_r^2 = \frac{(r\,!)^4}{(2r)!\ (2r+1)!}\ n(n^2-1)\,(n^2-2)\ \ldots\ldots\ (n^2-p)$$

(Kendall Vol.II)

$$= 0\ \frac{n^{2r+1}\ (r\,!)^4}{(2r)\,!\ \ (2r+1)\,!}$$

In particular

$$\sum^{n}\xi_1^2 = 0\left(\frac{n^3}{12}\right)$$

$$\sum^{n}\xi_2^2 = 0\left(\frac{n^5}{180}\right)$$

$$\sum^{n} \xi_3^{2} = 0\left(\frac{n^7}{2800} \right)$$

and $$\sum^{n} \xi_4^{2} = 0\left(\frac{n^9}{44100} \right) .$$

After simplification of algebra we obtain

$$\sum_{x=1}^{N=nk} \phi_1 = 0\left(\frac{nk}{2} \right)$$

$$\sum_{x=1}^{N} \phi_2 = 0\left(\frac{nk}{3} \right)$$

$$\sum_{x=1}^{N} \phi_3 = 0\left(\frac{n^3k}{20} \right)$$

$$\sum_{x=1}^{N} \phi_4 = 0\left(\frac{n^4}{8} - \frac{5n^3k}{168} \right)$$

Therefore

$$\frac{\sum_{x=1}^{N} \phi_1}{\sum^{n} \xi_1^{2}} = 0\left(\frac{6k}{n^2} \right)$$

$$\frac{\sum_{x=1}^{N} \phi_2}{\sum^{n} \xi_2^{2}} = 0\left(\frac{60k}{n^4} \right)$$

$$\frac{\sum_{x=1}^{N} \phi_3}{\sum^{n} \xi_3^{2}} = 0\left(\frac{140k}{n^4} \right)$$

and $$\frac{\sum_{x=1}^{N} \phi_4}{\sum^{n} \xi_4^{2}} = 0 \frac{44100}{n^6} \left(\frac{n}{8} - \frac{5k}{168} \right)$$

Thus the order of the upper limit of the bias

$$= \frac{6k}{n^2} \left| \sum^{n} \bar{y}_{\cdot j}\, \xi_1 \right| + \frac{60k}{n^4} \left| \sum^{n} \bar{y}_{\cdot j}\, \xi_2 \right| + \frac{140k}{n^4} \left| \sum^{n} \bar{y}_{\cdot j}\, \xi_3 \right| + \frac{44100}{n^{86}} \left(\frac{n}{8} - \frac{5k}{168} \right) \left| \sum^{n} y_{\cdot}\right.$$

4.11. It was not found possible to put the expression for the variance of T_p into a simple form so that in a direct or indirect manner the variance could be easily computed. The variance was completed, therefore, from the very definition and the efficiency of this estimate was compared with an arithmetic estimate based on simple random sampling.

4.12. A limited empirical study, regarding the bias present in the estimate T_q and the efficiency shown by it, was also made.

5. SAMPLING OF COWS AND OF THEIR DAILY MILK YIELDS FOR ESTIMATING AVERAGE PERFORMANCE OF A GROUP OF COWS.

5.1. In the previous section we have considered different methods of sampling and of estimating total lactation yields of individual animals. However, the problem of greatest practical interest is to obtain an objective estimate of the average lactation performance of a group of cows in a particular tract or under a given treatment. In the case where the number of cows is large it may not be possible to milk record all the cows if each cow is recorded only at periodic intervals and the question of sampling of a group of cows arises. It is in the light of variation between cows and between daily milk yields within cows as well as the desired level of precision that the number of cows to be sampled from a group of cows will have to be decided.

5.2. A scheme of sampling in which the selection of cows is made at random and recording of daily milk yield of selected cows done systematically at regular intervals would be convenient in practice.

5.3. Let M be the total number of cows

N_1 the lactation length of the 1-th cow

m the number of cows selected out of M cows.

k the interval of recording daily milk yields

n_1 the number of daily milk records kept for the 1-th cow.

Further let

y_{lij} denote the milk yield recorded on $(i + \overline{j-1}\,k)$-th day of the lactation of l-th cow,

$T_{l.}$ the actual lactation yield of the l-th cow

$$T_{l.} = \sum_{ij}^{N_l} y_{lij}$$

$\bar{T}_{..}$ the average lactation yield per cow in the populati

$$\bar{T}_{..} = \frac{1}{M} \sum_{l=1}^{M} T_{l.} = \frac{1}{M} \sum_{l=1}^{M} \sum_{ij}^{N_l} y_{lij}$$

$\bar{T}_{m.}$ the average lactation yield of a simple random sample of m cows

$$\bar{T}_{m.} = \frac{1}{M} \sum_{l}^{m} T_{l.}$$

T_{li} the total lactation yield of the l-th cow estimated from a systematic sample for which the first recording is made on the i-th day

$$T_{li} = k \sum_{j=1}^{n_l} y_{lij} .$$

This corresponds to T_{u} for one particular cow (c.f.4.4)

Now consider an estimate $\bar{T}_{m(k)}$ of the average lactation yield per cow in the population ($\bar{T}_{..}$) obtained from systematic sampling at intervals of k days of a random sample of m cows,

$$T_{m(k)} = \frac{1}{m} \sum_{l}^{m} T_{li} \tag{1}$$

$$= \frac{1}{m} \sum_{l}^{m} k \sum_{j=1}^{n_l} y_{lij}$$

5.3. Expected value of $\overline{T}_{m(k)}$

Since the sample is selected in two stages the expected value is also appropriately worked out in two stages, first over all possible systematic samples and then over all possible samples of m cows from the total of M cows . Thus we have

$$E(\overline{T}_{m(k)}) = E\left(\frac{1}{m}\sum_{l}^{m} T_{li}\right)$$

$$= E\left\{\frac{1}{m}\sum_{l}^{m} E(T_{li}/_{l})\right\}$$

where $E(T_{li}/l)$ denotes the expected value of T_{li} for a fixed l (i.e. for l-th cow)

It has already been shown that $E(T_{li}/l) = T_{l.}$

(T_u is an unbiased estimate of T Sec. 4.5)

$$\text{Therefore } E(\overline{T}_{m(k)}) = E\left\{\frac{1}{m}\sum_{l}^{m}(T_{l.})\right\}$$

$$= \frac{1}{M}\sum_{l=1}^{M} T_{l.} = \overline{T}_{..}$$

This shows that $\overline{T}_{m(k)}$ is an unbiased estimate of the average lactation yield in the population.

5.4. Variance of $\overline{T}_{m(k)}$

By definition the variance of the sample average lactation yield is given by

$$V(\overline{T}_{m(k)}) = E\left\{\overline{T}_{m(k)} - E(\overline{T}_{m(k)})\right\}^2$$

$$= E(\overline{T}_{m(k)} - \overline{T}_{..})^2$$

This can be written as

$$V(\overline{T}_{m(k)}) = E\,(\,\overline{T}_{m(k)} - \overline{T}_{m.} + \overline{T}_{m.} - \overline{T}_{..}\,)^2$$

$$= E(\overline{T}_{m(k)} - \overline{T}_{m.}\,)^2 + E(\overline{T}_{m.} - \overline{T}_{..})^2 + 2E\,(\overline{T}_{m(k)} - \overline{T}_{m.})(\overline{T}_{m.} - \overline{T}_{..}) \qquad (4)$$

Now $(\,\overline{T}_{m(k)} - \overline{T}_{m.}) = \frac{1}{m} \sum_{l}^{m} (T_{li} - T_{l.})$

So that

$$(\overline{T}_{m(k)} - \overline{T}_{m.})^2 = \frac{1}{m^2}\, E \left\{ \sum_{l}^{m} (\,T_{li} - T_{l.}) \right\}^2$$

$$= \frac{1}{m^2}\, E\left[\sum_{l}^{m} \left\{ (T_{li} - T_{l.})^2 \right\} + \sum_{l \neq l'} \left\{ (T_{li} - T_{l.})(T_{l'i'} - T_{l'.}) \right\} \right]$$

$$= \frac{1}{m^2} \left[E \sum_{l}^{m} E \left\{ (T_{li} - T_{l.})^2 / l \right\} \quad E \sum_{l \neq l'} E \left\{ (T_{li} - T_{l.})(T_{l'i'} - T_{l'.}) / l, i \right\} \right] \qquad (5)$$

The value of the second term under the summation sign is clearly zero since systematic samples are independently drawn from the l-th and l'-th cows and the value of the first term under the summation sign is given by expression for the variance of T_u derived in Section 4.6 .

$$E \left\{ (T_{li} - T_{l.})^2 / l \right\} = \frac{1}{k} \sum_{i=1}^{K} T_{li}^2 - T_{l.}^2$$

$$= \frac{1}{k} \sum_{i=1}^{K} (T_{li} - T_{l.})^2 = S_l^2(k) \text{ (say)} \qquad (6)$$

Substituting from (6) in (5) we thus obtain

$$E(\overline{T}_{m(k)} - \overline{T}_{m.})^2 = \frac{1}{m^2}\, E \sum_{l}^{m} S_l^2(k)$$

$$= \frac{1}{m} \cdot \frac{1}{M} \sum_{l}^{M} S_l^2(k)$$

$$= \frac{1}{m} \; \bar{S}^2_{.(k)} \tag{7}$$

where $$\bar{S}^2_{.(k)} = \frac{1}{M} \sum_{l=1}^{M} S^2_{l(k)} \tag{8}$$

It can be shown that second term in (4) is

$$E(\bar{T}_{m.} - \bar{T}_{..})^2 = \left(\frac{1}{m} - \frac{1}{M} \right) S_t^2 \qquad \text{(Sukhatme 1953)} \tag{9}$$

where $$S_t^2 = \sum_{l=1}^{M} (T_{l.} - \bar{T}_{..})^2/(M-1)$$

The value of last term in (4) is clearly zero. For

$$E \left\{ (\bar{T}_{m(k)} - \bar{T}_{m.}) (\bar{T}_{m.} - \bar{T}_{..}) \right\}$$

$$= E \left[(\bar{T}_{m(k)} - \bar{T}_{m.}) \times \frac{1}{m} \sum_{l=1}^{m} \left\{ (\bar{T}_{li} - \bar{T}_{l.})/1 \right\} \right]$$

$$= E \left\{ (\bar{T}_{m(k)} - \bar{T}_{m.}) \times 0 \right\} = 0 \tag{10}$$

Substituting (7), (9) and (10) in (4) we get on interchanging the order of the first two terms

$$V(\bar{T}_{m(k)}) = \left(\frac{1}{m} - \frac{1}{M} \right) S_t^2 + \frac{1}{m} \bar{S}^2_{.(k)} \tag{11}$$

5.5. Case I When complete daily milk records of a representative group of cows are available

a) Estimate of the variance of $\bar{T}_{m(k)}$:- The calculation of the variance of the sample mean in two-stage sampling involves the estimates of S_t^2 and $\bar{S}^2_{.(k)}$. The simplest way of estimating these is to define the corresponding quantities from the sample and take their expected values.

Let s_t^2 denote the mean square between totals of lactation yields obtained from cows in the sample

$$s_t^2 = \frac{1}{m-1} \sum_{l}^{m} (T_{l.} - \bar{T}_{m.})^2 \qquad (12)$$

and $\bar{s}^2_{\cdot(k)}$ denote the average of the variances of the estimated totals obtained from systematic samples recorded at k days interval, of individual cows in the sample

$$\bar{s}^2_{\cdot(k)} = \frac{1}{m} \sum_{l}^{m} \left\{ \frac{1}{k} \sum_{i=1}^{k} (T_{li} - T_{l.})^2 \right\} \qquad (13)$$

$$= \frac{1}{m} \sum_{l}^{m} s^2_{l(k)} \qquad (14)$$

when complete daily milk records of cows in the sample are available the statistics s_t^2 and $\bar{s}^2_{\cdot(k)}$ are unbiased estimates of the corresponding population values S_t^2 and $\bar{S}^2_{\cdot(k)}$.

Thus we get the estimate of the variance of $\bar{T}_{m(k)}$ in the present case.

$$\text{Est } V(\bar{T}_{m(k)}) = \left(\frac{1}{m} - \frac{1}{M}\right) s_t^2 + \frac{1}{m} \bar{s}^2_{\cdot(k)} \qquad (15)$$

b) Determination of sample size :- The expression (15) of the estimate of variance of $\bar{T}_{m(k)}$ can be utilized for estimating the precision that may be expected to be attained for any given scale of sampling (of cows and of daily yields) following the pattern of sampling and of estimation given at (5.2) . Conversely the basic data and the expression can be used for working out the number of cows required to be sampled for milk recording at specific intervals for estimating the average lactation yield of a group of cows with prescribed level of precision. If p is the percentage sampling error with which it is desired to estimate the average lactation yield of the group then the number of cows required can be worked out from the following formula.

$$\frac{S.E.(\bar{T}_{m(k)})}{\bar{T}_{m(k)}} \times 100 = p \qquad (16)$$

Squaring both sides, substituting the estimate of variance of $\bar{T}_{m(k)}$ and rearranging the terms we get

$$\left(\frac{1}{m} - \frac{1}{M}\right) s_t^2 + \frac{\bar{s}_i^2(k)}{m} = \frac{p^2 \bar{T}_{m(k)}^2}{(100)^2}$$

or

$$m = \frac{(s_t^2 + \bar{s}_i^2(k))(100)^2}{p^2 \bar{T}_{m(k)}^2 + (100)^2 \frac{s_t^2}{M}} \qquad (17)$$

for very large M we get

$$m = \frac{(s_t^2 + \bar{s}_i^2(k))(100)^2}{p^2 \bar{T}_{m(k)}^2} \qquad (18)$$

5.6. Case II Estimate of the variance of $\bar{T}_{m(k)}$ when one systematic sample for each of the selected cow is available and M is very large.

The expression of the variance of the estimate given in Section 5.4 cannot be directly used to obtain the estimate of the standard error of the average lactation yield in the case when only one systematic sample is available .

since the values of S_t^2 and $S^2_{(k)}$ cannot be estimated. However, an approximate estimate of the variance is developed in this section.

Let s_{ti}^2 denote the mean square between totals of lactation yields obtained from cows in the sample.

$$s_{ti}^2 = \frac{1}{m-1} \sum_{\ell}^{m} (T_{li} - \bar{T}_{m(k)})^2 \tag{19}$$

or
$$(m-1)\, s_{ti}^2 = \sum_{\ell}^{m} (T_{li} - \bar{T}_{m(k)})^2 \tag{20}$$

$$= \sum_{\ell}^{m} T_{li}^2 - m\, \bar{T}^2_{m(k)} \tag{21}$$

Taking expectations we get

$$(m-1)\, E(s_{ti}^2) = E \sum_{\ell}^{m} T_{li}^2 - m\, E (\bar{T}^2_{m(k)}) \tag{22}$$

Now
$$E(T_{li}^2) = V(T_{li}) + E^2 (T_{li}) \tag{23}$$

The $V(T_{li})$ is given from the expression (11) by putting $m = 1$

Therefore
$$V(T_{li}) = (1 - \frac{1}{M})\, S_t^2 + \bar{S}^2_{i(K)} \tag{24}$$

and
$$E^2(T_{li}) = E^2 \left\{ E(T_{li}/l) \right\} = E^2 (T_{l.}) = \bar{T}^2_{..} \tag{25}$$

Now to obtain $E\, (\bar{T}_{m(k)})^2$ we can write it as

$$E(\bar{T}_{m(k)})^2 = V(\bar{T}_{m(k)}) + E^2 (\bar{T}_{m(k)})$$

which from (2) and (11) can be expressed as

$$(\frac{1}{m} - \frac{1}{M})\, S_t^2 + \frac{1}{m}\, \bar{S}^2_{i(k)} + \bar{T}^2_{..} \tag{26}$$

Thus $E\, (s_{ti}^2)$ with the help of (18), (2[illegible]), (24), (25) and (26) is given by

$$\frac{m}{m-1}\left\{\left(1-\frac{1}{M}\right)S_t^2+\bar{S}^2_{\cdot(k)}+\bar{T}^2_{\cdot\cdot}-\left(\frac{1}{m}-\frac{1}{M}\right)S_t^2-\frac{1}{m}\bar{S}^2_{\cdot(k)}-\bar{T}^2_{\cdot\cdot}\right\}$$

$$=\frac{m}{m-1}\left\{\frac{m-1}{m}S_t^2+\frac{m-1}{m}\bar{S}^2_{\cdot(k)}\right\}$$

Therefore $E(s^2_{t1}) = (S_t^2 + \bar{S}^2_{\cdot(k)})$ (27)

Thus s^2_{t1} is an unbiased estimate of $(S_t^2 + \bar{S}^2_{\cdot(k)})$ Hence the estimate of the $V(\bar{T}_{m(k)})$ in case when M is very large is given by

Est. $V(\bar{T}_{m(k)}) = \left[\frac{1}{m}-\frac{1}{M}\right]s^2_{t1} = \left(\frac{1}{m}-\frac{1}{M}\right)s^2_{t1}$ (28)

It can be seen from (15) that the estimate is slightly an underestimate in so far as it ignores the term $(\frac{1}{M}\bar{S}^2_{\cdot(k)})$

5.7. <u>Case III</u> Estimate of the variance of $\bar{T}_{m(k)}$ when only one systematic sample for each cow <u>of the selected cows is available and M is finite.</u>

In general case, however, when M is finite and not large in size it will be seen from Sec. 5.6 that we cannot estimate the variance of the estimate unless an estimate of $\bar{S}^2_{\cdot(k)}$ or of another independent linear equation in S_t^2 and $\bar{S}^2_{\cdot(k)}$ is available. But it is well known that for systematic sampling it is not possible to estimate the variance of a single lactation yield ($S^2_{1(k)}$) from a sample. However, with a'priori knowledge regarding the efficiency of systematic sampling over simple random sampling or regarding the intraclass correlation (since the efficiency and the correlation are linearly related) obtained from some previous studies on complete enumeration data, the difficulty of estimating $\overline{S}^2_{.(k)}$ can be overcome. We can write

$$S^2_{l(k)} = N_l^2 \left(1 - \frac{1}{N_l} \right) \frac{S^2_{lo}}{n_l} \left(1 + \rho_l \overline{n_l - 1} \right) \tag{29}$$

where $S^2_{lo} = \frac{1}{N_l - 1} \sum_{j=1}^{N_l} (y_{lij} - \bar{y}_{l.})^2$ in which

$\bar{y}_{l.}$ stands for $\frac{1}{N_l} \sum_{j=1}^{N_l} y_{lij}$

and ρ_l the intraclass correlation between daily milk yields within k-columns of n_l records each, which can be formed out of N_l records of the l-th cow .

Thus the expression (11)

$$V(\overline{T}_{m(k)}) = \left(\frac{1}{m} - \frac{1}{M} \right) \{ S_t^2 + \overline{S}^2_{.(k)}) + \frac{1}{M} \overline{S}^2_{.(k)})$$

can be written as

$$V(\overline{T}_{m(k)}) = \left(\frac{1}{m} - \frac{1}{M} \right) (S_t^2 + \overline{S}^2_{.(k)}) + \frac{1}{M^2} \sum_{l=1}^{M} N_l (N_l - 1) \frac{S^2_{lo}}{n_l} \left(1 + \rho_l \overline{n_l - 1} \right) \tag{30}$$

It is well known that s^2_{lo} is an unbiased estimate of S^2_{lo}

where $s^2_{lo} = \frac{1}{n_l - 1} \sum_{j=1}^{n_l} (y_{lij} - \bar{y}_{ln_l})^2$

and $\bar{y}_{ln_l} = \frac{1}{n_l} \sum_{j}^{n_l} y_{lij}$

It can be easily shown that

$$\frac{1}{mM} \sum_{l}^{m} N_l (N_l - 1) \frac{s_{lo}^2}{n_l} (1 + \rho_l (n_l - 1))$$

is an unbiased estimate of the second term in the expression (30) provided the knowledge regarding is a'priori and it has already been shown in the previous section that $E (s_{t1}^2) = S_t^2 + \bar{S}^2_{.(k)}$

$$\text{thus Est } V(\bar{T}_{m(k)}) = \left(\frac{1}{m} - \frac{1}{M}\right) s_{t1}^2 + \frac{1}{mM} \sum_{l}^{m} N_l (N_l - 1) \frac{s_{lo}^2}{n_l} (1 + \rho_l (n_l - 1)) \qquad (31)$$

5.8. We have seen in section 4.4 that there was an alternative estimate of the individual lactation yield based on average per day which was slightly biased but was more efficient as compared to the estimate based on sample total, for monthly or larger intervals(sec. 6.3). Corresponding to it we can develop an estimate $\bar{T}_{m_b}(k)$ of the average performance of a group of cows

$$\bar{T}_{mb(k)} = \frac{1}{m} \sum_{l}^{m} \frac{N_l}{n_l} \sum_{j=1}^{n_l} y_{lij}$$

This is a biased estimate but the bias is negligible being the average of all the values of biases for an estimate T_b of individual cows which as we have seen (sec.6.1) are themselves small.

The term $\bar{S}^2_{.(k)}$ in the expression (11) of the variance of $\bar{T}_{m(k)}$ will be replaced by another similar term $\bar{s}^2_{\cdot b}(k)$ corresponding to the variances of T_b for getting the variance of $\bar{T}_{m_b}(k)$ in case the bias present in $\bar{T}_{m_b}(k)$ is ignored.

The scheme of sampling and estimating described in sections 5.2 and 5.3 could have been compared with any other scheme of two-stage sampling in which the selection of cows at the first stage is made at random but at the second stage recording is either (i) made at random and production estimated by arithmetic estimate or (ii) or done at regular intervals and the total lactation yield is estimated taking due account of the knowledge regarding the trend of the lactation curve. This was not, however, considered necessary since the variance of any estimate of average lactation yield from either of the above mentioned schemes depends overwhelmingly upon the contribution from the term containing variation between lactation yields i.e. s_t^2 .

6. RESULTS

6.1. At the outset the extent of bias in the estimate T_b was worked out for each of the 41 cows at four intervals of recording viz., 7-day, 14-day, 28-day and 56-day intervals.

For weekly interval of recording the maximum values of the bias was found to be 0.02 per cent which in a herd of lactation yield of 1200 pounds will amount to 4 ounces. The values lay in the interval of -0.06 to 0.06 per cent for recording at biweekly intervals. For 28-day and 56-day intervals of recording the frequency distribution of the percentage biases has been shown in the following table :-

Table I.

Table showing the frequency distribution of percentage biases in the estimate T_b for different intervals of recording.

28-day interval

Percentage Bias.	Number of cows.
-0.10-	2
0 -	30
0.10-	8
0.20-	1
	41

56-day interval

Percentage Bias.	Number of cows.
-0.30-	2
0 -	22
0.20-	10
0.40-	3
0.60-	3
0.80-	1
	41

The occurrence of negative bias is very rare. This indicates that the estimate T_b of the total lactation yield overestimates the actual yield. However, this overestimation is never more than one per cent. The majority of the biases are less than 0.2 and 0.4 per cent for 28-day and 56-day of recording which in herd of a lactation yield of 1200 pounds would amount to 2.4 to 4.8 pounds.

6.2. Standard errors of the estimates T_u and T_b, both based on systematic samples, expressed as percentage of actual lactation yield were next calculated for the four intervals of recording. Table II shows the frequency distribution of the percentage standard error of these two estimates and of another estimate T_r obtained from simple random samples of corresponding sizes for the four intervals of recording.

Table II.

Distribution of percentage standard errors for each of the estimates (T_r, T_u and T_b) for different intervals of recording.

Percentage standard error.	7-day interval.			14-day interval			28-day interval			56-day interval		
	T_b	T_u	T_r	T_b	T_u	T_r	T_b	T_u	T_r	T_b	T_u	T_r
less than one.	7	9										
1-	25	23		10	13							
2-	7	7	12	20	15		2	2				
3-	2	2	22	10	9	3	16	10				
4-			4	1	2	20	14	17		3		
5-			3			12	4	4	8	10	4	
6-					2	1	3	3	10	10	5	
7-						4	2	3	13	5	10	2
8-						1		1	4	3	7	8
9-									2	4	7	9
10-								1	2	2	0	10
11-									1	2	5	5
12-									1		0	2
13-											2	0
14-												2
15-												1
16-												1
17-											1	1
18-										1		
19-										1		
20-												
Total:	41	41	41	41	41	41	41	41	41	41	41	41

The majority of the standard errors for the estimates based on weekly recordings were observed to fall in the interval of 1 to 2 per cent. The average standard error was about 1.5 per cent. The average standard error for the corresponding estimate obtained from simple random samples was 3.4 per cent.

The standard errors of the estimate T_u for 14-day interval of recording were always less than 5 per cent with only two exceptions, which correspond to the estimates obtained from the data of two cows which were poor yielders. The high percentage standard errors in these cases may be partially ascribable to this cause. For the same interval of recording the standard error of the estimate T_b were never greater than 5 per cent. On an average the standard error for the estimates T_b and T_u were 2.5 and 2.6 per cent respectively. The standard error of the estimate T_r, based on simple random samples of the corresponding sizes were more than 5 per cent in no less than 43 per cent of the cases.

For 28-day interval of recording, leaving aside a few of the standard errors of the two estimates, T_b and T_u, obtained from cows with smaller lactation lengths, were rarely more than 5 per cent. On an average these were 4.4 and 4.9 per cent for these estimates. However, for the corresponding random sampling estimate T_r the standard errors were never less than 5 per cent and averaged out to 7.4 per cent.

For the 56-day interval of recording the standard errors turned out to be 7.7, 8.5 and 10.6 per cent on average basis for the three estimates T_b, T_u and T_r respectively.

The frequency distributions clearly show that the estimate T_b tends to have smaller standard error as compared to T_u. It is obvious that the means will be more stable as compared to total of milk yields from samples of relatively different sizes from one another.

It will be seen from the results that in case the estimates of the lactation yield are desired to be obtained with a standard error of less than 5 per cent on an average, the milk production should be calculated by multiplying the average yield obtained from records made systematically at monthly or larger intervals.

6.3. The percentage gains in efficiency of the two arithmetic estimates viz., T_u and T_b of the total lactation yield from systematic samples using recording intervals of 7, 14, 28 and 56 days over similar procedures of estimation from simple random samples of same sizes were obtained next. If S is the variance of the estimates (T_u or T_b) of the total lactation yield based on systematic sample for a given interval and R is the variance of estimate based on random samples of same size then percentage gain in efficiency of estimates based on systematic samples over those based on simple random samples is taken to be $(\frac{R}{S} - 1)\,100$.

The values of S in the two cases - T_u and T_b - were obtained, by employing the formulae (Section 4.6), by regarding the daily yields of each cow as constituting a finite population. The value of R is obtained from the formula

$$R = \left(\frac{1}{n} - \frac{1}{N}\right) \frac{\sum_{ij}^{N} y_{ij}^2 - \left(\sum_{ij}^{N} y_{ij}\right)^2 / N}{N-1}$$

where N is the lactation period of a cow

n the size of the random sample

y_{ij} the milk yield recorded on $(i + \overline{j-1}\, k)$th day

The values of the percentage gains inefficiency of the estimates (T_u and T_b) for each of the 41 cows are given in Table III.

TABLE III

Percentage gains in efficiency of two procedures of estimation of total lactation yield from systematic samples over similar procedure of estimation from simple random samples.

S.No. of cows.	7-day interval.		14-day interval		28-day interval		56-day interval	
	T_u	T_b	T_u	T_b	T_u	T_b	T_u	T_b
1	234	225	55	282	14	181	-18	-30
2	452	75	-11	215	-1	267	18	68
3	345	1333	668	573	209	780	83	540
4	1719	991	441	383	382	206	212	115
5	521	708	784	1346	86	415	243	90
6	53	143	18	145	68	206	10	16
7	375	375	328	328	415	415	185	185
8	445	1759	430	1070	165	6	107	- 67
9	1261	764	1028	300	590	200	99	113
10	183	118	146	147	39	338	- 9	124
11	1622	1622	447	447	18	342	59	350
12	- 62	- 62	574	574	119	184	102	-10
13	377	377	194	194	155	196	93	261
14	829	317	201	234	239	267	71	87
15	1012	762	393	240	139	172	361	505
16	452	120	240	439	57	206	29	-65
17	167	264	12	367	-16	-16	-27	-27
18	475	475	215	127	197	1	0	-27
19	500	1137	16	147	16	315	-24	429
20	19	51	28	134	15	-4	-18	-21
21	3227	1296	858	1158	622	328	192	234
22	1385	540	1064	143	149	152	9	226
23	287	36	393	317	99	107	7	131
24	564	270	336	226	300	203	186	70
25	527	697	106	290	141	407	-48	127
26	81	81	197	197	79	79	21	117
27	749	749	643	643	181	181	18	133
28	1165	1165	1129	1129	470	470	30	589
29	212	212	219	219	215	215	29	225
30	592	276	481	559	38	60	-25	33
31	1440	1440	169	343	39	450	50	179
32	1343	1343	1218	672	203	131	61	160
33	575	575	379	560	312	465	305	519
34	268	268	290	290	274	274	243	243
35	242	111	213	72	151	162	148	110
36	124	124	804	804	420	438	351	266
37	-33	-33	356	434	136	475	61	420
38	135	135	19	206	77	88	84	34
39	818	818	680	680	122	122	37	143
40	323	279	170	159	343	79	351	266
41	618	618	780	269	118	121	49	42
Average	625	579	408	441	180	236	91	168

T_u is (unbiased) arithmetic estimate based on systematic sample total.

T_b (biased) arithmetic estimate based on systematic sample mean.

From the table it is seen that for recording at weekly intervals the average gain in efficiency of the estimate T_u was 625 per cent. For biweekly weighings the estimate T_u had always shown gain in efficiency with a single exception (Cow No.2). The average gain in efficiency was 406 per cent. In 29 per cent of the cases the values of the gains were more than 500 per cent.

For four weekly interval of recording the estimate T_u had shown losses in efficiency in nearly 5 per cent of the cases. 66 per cent of the gains were less than 200 per cent. The average gain in efficiency for this interval of recording was 180 per cent- slightly less than half the gain obtained for estimate T_u based on biweekly weighings. Only 12 per cent gains exceeded 400 per cent.

For recording interval of 56 days in as many as 17 per cent of the cases the losses in efficiency of the estimate T_u had occurred. All gains in efficiency were less than 400 per cent and the majority of them lay below 200 per cent. The average gain in efficiency was 91 per cent. This was half of the gain in efficiency of the estimate obtained from 28-day interval of recording.

Coming to the efficiency of the estimate T_b based on weekly weighings it may be seen that the average gain in efficiency was 579 per cent. This was less than the corresponding gain observed for the estimate T_u.

For biweekly weighings the distribution of the gains in efficiency of the estimate T_b was almost similar to that of the gains in efficiency of T_u at the interval of recording. The average gain in efficiency was 441 per cent.

For recording interval of 28 days more than 50 per cent of the gains were seen to be between 200 and 500 per cent. As in the case of the estimate T_u the percentage losses in efficiency had occurred in nearly 5 per cent of the cases for the estimate T_b. The average gain in efficiency was 236 per cent.

In case of eight weekly interval of recording, the losses in efficiency for T_b were as frequent as in the case of the estimate T_u for this interval of recording. The average gain in efficiency came out to be 168 per cent.

But for weekly interval of recording, the estimate T_b was more efficient as compared to the estimate T_u.

6.4. It was anticipated that the estimates T_q and T_p got from systematic samples by utilising the knowledge regarding the lactation curves would be more efficient, though biased, as compared to arithmetic estimates obtained from systematic samples. To calculate the bias and efficiency at 14-day interval of recording even from the data of one cow as many as 28 curves would have to be fitted. This number of curves becomes double and fourfold for calculation of bias and efficiency at 28-day and 56-day intervals of recording respectively. In view of the heavy and extremely voluminous calculation involved, it was not possible, in course of the time alloted for this thesis, to take all cows at all intervals.

However, in order to obtain a rough idea of the order of magnitude of bias and of the further gain in efficiency by utilisation of the trend of the curve eight cows having lactation lengths round about the average lactation length of 301 days were selected and curves fitted to find both bias and efficiency of T_q and T_p. Of the selected cows one (No. 10) had a lactation period of 293 days, three (cow Nos. 11, 12 and 13) with lactation period of 294 days each and the rest four (cow Nos. 26, 27, 28 and 29) had all a lactation period of 308 days. These cows were chosen with a view to minimising the computational work, since a set of fourteen inverted matrices and the fourteen sets of values of $\sum_{x=1}^{N} \xi_1$, $\sum_{x=1}^{N} \xi_2$, $\sum_{x=1}^{N} \xi_3$ and $\sum_{x=1}^{N} \xi_4$ computed for a cow would remain invarient for other cows of the same lactation period.

a) <u>Bias</u> :- The expression for the upper limit of the bias in T_p has been obtained in Section 4.10. To get the order of $\sum_{j}^{n} \bar{y}_{.j} \xi_r$ the values of $\bar{y}_{.j}$ were taken from the lactation curve for a group of cows having lactation length near about 300 days. The equation of the lactation curve (Khandekar was taken to be

$$y = 2.137 + 0.0652x - 0.0006232x^2 + 0.000{,}002117x^3 - 0.000{,}000{,}002269x$$

where y stands for milk yield in pounds and x the advance in lactation in days.

Taking 14-day interval of recording and assuming the lactation length to be of 294 days, we get, on substituting numerical values for algebriac terms in the upper limit of the bias.

Thus the upper limit

$$\begin{aligned} &= (.1905)(2.79) + (.00432)(691.81) + (0.017)(1280.99) \\ &\qquad + (0.001135)(1465.04) \\ &= 0.531 + 2.898 + 12.910 + 1.663 \\ &= 18.002 \end{aligned}$$

We have already seen that average lactation yield of the cows is 1201 pounds. Thus the upper limit of the bias is 1.5 per cent.

The bias in T_q and T_p were also empirically calculated from the daily milk yields of the above mentioned eight cows. These were all less than one per cent, but for the bias in estimate T_q for cow No. 11, in which case the value turned out to be 1.16 per cent. Since there were negative and positive values of bias, simple arithmetic mean would not have given an idea of average bias. Absolute averages were worked out. These came out to 0.42 and 0.11 per cent in estimates T_q and T_p respectively. This was in consonance with the theoretical result that the upper limit of bias in T_p is 1.5 per cent.

b) <u>Efficiency</u> : The gains in efficiency of the estimates T_q and T_p over arithmetic estimate from simple random samples are shown, side by side with the gains in efficiency of the arithmetic estimate (T_u) from systematic samples, in table IV. All the gains are expressed in percentages.

Table IV

Table showing the percentage gains in efficiency of the estimates T_q, T_p and T_u from systematic samples at 14-day interval of recording over T_R.

Cow No.	Percentage gain in efficiency of the estimates :- T_u	T_q	T_p
10	146	469	389
11	447	808	1031
12	574	311	510
13	194	312	226
26	197	290	275
27	643	706	902
28	1129	1515	1244
29	219	278	269
Average:	444	586	606
Average excluding cow No.12.	425	625	620

Results in Table IV indicate that with the exception of only one case i.e. cow No.12 the gain in efficiency of the estimates T_q and T_p is always more than that of T_u. On an average the gains in efficiency of T_q and T_p are 142 and 162 per cent more as compared to T_u. Excluding cow No.12 the gains, in efficiency of T_p and T_q are on an average nearly 200 per cent more as compared to the gain in efficiency of T_u.

6.5 Turning next to the problem of estimating average lactation yield of a group of cows the estimate of variance between lactation yields of cows (s_t^2) was found to be 108148.72 (pounds)^2 . The variances of the estimated lactation yields obtained from systematic samples ($\bar{s}^2_{.:k}$)) for 7-day, 14-day 28-day and 56-day intervals of recording were estimated as 405.84, 948.46, 3226.61 and 11369.41(pounds)^2 respectively. With these components of variance, the number of cows required for estimating the average lactation yield of groups of cows of varying strength by sampling at 7-day, 14-day, 28-day and 56-day intervals for different levels of precision aimed at, were worked out. These are given in table below :-

Table V

Table showing the number of cows required for estimating the average lactation yield of a groups of cows of different sizes and different levels of precision for the four intervals of recording.

Herd size.	Interval of recording.	Level of precision(Percentage standard of the estimate.) 1	2	3	5	10
100	7-day	69	56	47	24	8
	14-day	90	66	48	24	8
	28-day	91	68	47	24	8
	56-day	98	73	51	26	8
300	7-day	216	116	66	28	8
	14-day	217	117	66	28	8
	28-day	221	119	68	29	8
	56-day	236	128	73	31	9
500	7-day	302	138	73	29	8
	14-day	303	138	73	29	8
	28-day	309	141	74	30	8
	56-day	332	151	80	32	9
2000	7-day	549	172	81	30	8
	14-day	551	173	81	30	8
	28-day	562	177	83	31	8
	56-day	603	190	89	33	9
5000	7-day	653	181	82	30	8
	14-day	657	182	83	30	8
	28-day	671	186	84	31	8
	56-day	720	200	91	33	9
∞	7-day	754	189	85	31	8
	14-day	757	190	84	31	8
	28-day	773	194	86	31	8
	56-day	829	208	93	34	9

The table V gives the number of cows to be sampled for estimating at prescribed levels of percentage standard error, the average lactation yield of herds of different sizes by sampling at given intervals. It will be seen from the results that the number of cows to be sampled is almost the same for a given herd size and level of precision whatever be the interval of recording, except for a small increase for 8-weekly recording for low value of standard error. This is to be ascribed to the high magnitude of the mean square between lactation yield of cows compared to the within cow variation. Further it is seen that average lactation yield may be estimated with about 5 per cent standard error by sampling about 25 cows from small herd of size 100 and about 30 to 35 cows from herds of larger size.

6.6. As regards the precision of the estimation of average lactation yield, the variance of the estimate can be obtained approximately by following the procedure in Sec.5.6. As shown there the variance will be slightly underestimated. The extent of underestimation was worked out with the present data for the different scales of sampling and herd size considered in section 6.5. The extent of underestimation was extremely negligible for cases in which the estimates were xpected to be obtained with standard errors of the order of 10 per cent except for herds of very small size, such as 100, in which for 56-day of sampling the underestimation approached one per cent. For scale of sampling designed to provide very precise estimates of the order one per cent the underestimation was much greater especially for larger intervals of sampling, the value being as much as 90 per cent for 56-day interval when the herd size was as small as 100. The value reduced to 30 per cent for herd size 300 for the same interval. For larger herd sizes even with 56-day interval the underestimation was less than 2 per cent.

7. DISCUSSION AND CONCLUSIONS

It is clear from the results that the arithmetic estimates based on systematic are far more efficient than the estimate based on simple random sampling apart from being convenient in practice.

Recording of milk yield at weekly intervals would provide an estimate of a single lactation yield with a precision corresponding to a standard error of about 1.5 per cent, which should be considered reasonably satisfactory for the purpose of rational supervision over daily milk recording by the research staff on a farm. An estimate based on weekly recording cannot, however, be relied upon in cases where the lactation yield is required to be estimated with a margin of less than 3 percent.

In the light of the fact that the bias present in T_b is negligible and because the increase in efficiency over T_u increases with the interval of recording, it may be recommended that the estimate T_b may be adopted for monthly or larger intervals of recording.

It can be seen from table IV that further gain of about 150 per cent is indicated when the estimate of a single lactation yield is obtained by utilizing the knowledge regarding the lactation curve. However, in view of the small number of cows that has been analysed for this item of the study, it is not possible to make a firm recommendation.

As far as bringing further gain in efficiency is concerned, there is nothing to choose between the two curves, but from the point of view of fitting and estimation the polynomial curve is more convenient, easy and accurate especially when the method described in section 4.9 is adopted.

While reviewing the literature it was found that many of the workers have obtained the percentage error of estimating lactation yield on the basis of recording made at various intervals. For the sake of comparison similar errors were computed. The maximum and minimum errors were of the order of 10 and -10; 22 and -26 ; and 35 and -30 respectively for biweekly, four weekly and eight weekly intervals of recording. These values are consistently higher than those found by other authors, for instance Jordao et al obtained the values of 7 and -7; 12 and -11 ; and 12 and -14 for the same intervals. It was observed that approximately 80 per cent of the errors were lying in the interval -3 to 3 for 14-day interval, -6 to 6 for 28-day interval and -12 to 12 for 56-day interval. On an average the errors of order 2.1, 4.4 and 7.6 were observed for the three intervals of recording. These also are higher as compared to those obtained by others, e.g., McDowell who had values of the order of 2.91 for monthly and 3.80 for bimonthly recording.

The results regarding the number of cows required to be sampled for estimating the average lactation yield of a group of cows can be utilised in assessing the merit of the alternative plans for milk recording in the key village scheme. A Key Village block consisting of six key villages contains about 5000 cows of breeding age. It is envisaged that a milk recorder can record the yields of atleast three selected cows in milk, morning and evening, every working day. He can, therefore, be expected to record the yields of 75 cows at monthly intervals or 150 cows if recorded at two monthly intervals. For these two alternative systems of recording the percentage standard errors with which the average lactation yield may be expected to be estimated were worked out from the data. It was observed that the value was 3.2 per cent with monthly recording while it was reduced to 2.3 per cent with recording at two-monthly intervals. It is clear that in cases when the object is to estimate the average lactation yield of a group of cows and not to assess the performance of individual cows two-monthly recording is to be preferred.

During the year of investigation the study could only be carried upto the stage indicated in the foregoing paragraphs. The investigations need to be pursued further and some of the outstanding problems are as follows :-

The conclusions drawn from the present study are to be taken as only tentative in the sense that they are based only on a limited study of 41 cows belonging to a single herd. It would be desirable to study the data on other herds of various breeds as also those on village cattle to see how far the conclusions are of a general application. A limited empirical investigation has been conducted in the present thesis on the possibility of increasing the efficiency of the estimate by utilising the knowledge regarding the trend of the lactation curve. This needs to be followed up and if the line proves to be promising, a simple and practicable procedure for working out the estimate from systematically recorded data will have to be developed. A theoretical procedure for an indirect estimation of the variance of an estimate based on systematic sampling and utilising the knowledge of the curve also needs to be evolved. In the case of individual cows on farms the possibility of increasing the precision by sampling at varying intervals in different stages of lactation needs to be examined. Yates (1948) has suggested arbitrary end-corrections to the estimates obtained from systematic sampling for artificially reducing the variability between the estimates from samples with different starting observations. The applicability of the corrections proposed by him or the suitability of any modified correction has also to be examined.

8. SUMMARY

1. For assessing the progress of cattle development schemes like the key village scheme, which operate on village cattle, it is necessary to estimate objectively the average lactation yield of the animals in the scheme. Maintenance of daily milk records being prohibitive in cost, it is necessary to evolve a suitable procedure of sampling of a group of cows as also the days in milk of the selected cows and to develop an efficient method of estimation.

2. As a first approach to the problem, the daily milk records of 41 Hariana cows collected at the Government Livestock Farm, Hissar have been examined to study how far a procedure of sampling can be recommended.

3. It was found that the average lactation period of the herd was 301 days. The average lactation yield and the average daily milk yield were 1201 pounds and 3.9 pounds respectively. The coefficient of the variation between daily yields was 25 per cent.

4. The recording of milk yield at weekly intervals would provide an estimate of a single lactation yield with a precision corresponding to a standard error of about 1.5 per cent, which would be considered reasonably satisfactory for the purpose of rational supervision. An estimate based on weekly recording cannot, however, be relied upon in cases where the lactation yield is required to be estimated with a margin of error less than 3 per cent.

5. The bias in the estimate T_b, of a single lactation yield, based on systematic sample mean was found to be less than one per cent for all the intervals of recording considered.

6. The gains in efficiency of the two arithmetic estimates, T_u and T_b, former based on the systematic sample totals and latter on systematic sample means, were 625, 408, 180 and 91; and 579, 441, 236 and 168 respectively, for recording at one week, two weeks, four weeks and eight weeks.

In the light of the fact that the bias present in T_b is negligible and because the increase in efficiency over T_u increases with the interval of recording, it may be recommended that the estimate T_b may be adopted for monthly or larger intervals of recording.

7. A limited empirical study indicated that there is a further gain in efficiency of about 150 per cent when the knowledge regarding the lactation curve is utilised in estimating a lactation yield from systematic sample recorded at 14-day interval. However, in view of the small number of cows that has been analysed for this item of study, it is not possible to make a firm recommendation.

8. From the mean squares, between the lactation yields of the cows and between systematic sample estimates within cows the number of cows required to be selected for milk recording at given intervals for estimating the average lactation yield of herds of different sizes with percentage standard errors of prescribed magnitude was found. The number of cows required to be sampled was almost the same for a given herd size and

level of precision irrespective of the interval of recording. It was found that the average lactation yield may be estimated with 5 per cent standard error by sampling about 25 cows from a small herd of size 100 and about 30 to 35 cows from herds of larger size.

9. A method of approximately estimating the variance of the estimated average lactation yield was developed. It was shown that the estimate gives a slight underestimation of the variance which is less than 2 per cent for even bimonthly sampling in herds of size larger than 300. However, for smaller herds the underestimation is likely to be higher for longer intervals of sampling.

10. With the results obtained, the plan for milk recording in a village development programme such as the key village scheme for the purpose of estimating the average lactation yield of the population in the scheme was examined. A key village block consisting of six key villages contains about 5,000 cows of breeding age. It is envisaged that a milk recorder can record yield of three cows morning and evening every working day. He can thus record the yields of about 75 cows at monthly intervals or of 150 cows at bimonthly intervals. It was observed that the percentage standard error of estimate of the average lactation yield of the population of breeding cows was 3.2 per cent with monthly recording and 2.3 for bimonthly recording. It seems that in cases when the object is to estimate the average lactation yield of a group cows and not to assess the performance of individual cows bimonthly recording is to be preferred.

-----

9. REFERENCES

*1. Ashton, E.D.(1956) "Milk and Butterfat recording."

Tech. Communication No.3
Commonwealth Bureau of Dairy Sci.& Tech.
Commonwealth Agri. Bureaux
Farnham Royal Bucks , England.

2. Cannon,C.Y., Frye,J.B., and Sims, R.(1942) "Predicting 305-day yields from short-time records"

J. Dairy Sci. 25(12) 991-99.

*3. Cochran, W.G.(1946) "Relative accuracy of systematic and stratified random samples for a certain class of populations."

Ann. Math. Stat. Vol.XVII pp.164-77

4. Copeland, L. (1928) "Monthly and Bi-monthly Tests".

Jersey Bull. 47 pp. 731-32.

*5. De Lury B.D.(1950) "Values and Integrals of the Orthogonal Polynomials upto n = 26."

Toronto. University of Toronto Press.

6. Dick, I.D.(1950) "Some results on the accuracy of milk sampling over a complete lactation and the daily variability in milk-yield."

New Zealand J. Sci. & Tech.
Vol.32, Sect. A(1), 25-29,(

7. Erb, R.E., Goodwin, M.M., Morrison, R.A. and Shaw,A.O. (1952). "Lactation studies. III. Variations in daily yields of dairy cows. IV. Accuracy of different methods of estimating lactation yields."

J. Dairy Sci. 35: 967- 76, 977-87.

8. Flishmann W.(1891) "Investigations on the milk from 16 pure bred Friesian Cows in East Prusia throughout one lactation."

Landw Jb. 20; Ergänzungsband 2

9. Fisher,R.A.(1950) "Statistical methods for Research workers.

Oliver and Boyd , England.

---

* References not cited in the body of the thesis .

10. Fisher, R.A and Yates, F. (1953) "Statistical Tables for Biological, Agricultural and Medical Research." Oliver and Boyd.

11. Gaines, W.L., (1927) "The deffered Short-time test as a measure of the performance of dairy cows." J. Agric. Res. 35(3)

12. Houston, J and Hale, R.W. (1932) "The errors involved in certain methods of estimating the lactation yield of milk and butterfat." J. Dairy Res. 4(1) 37-47

13. Jordim, W.R, Peixoto, A.M., Filho, S.S. and Gomes, F.P. (1956) "Study on the precision and accuracy of a few practical methods of estimating milk production of dairy cows. Biometrics; 12(2): Abstracts (333)

14. Kendall M.G. (1946) "The advanced theory of Statistics." Vol.II- pp-159-62 Charles Griffin & Co. Limited, London.

15. Khandekar, N.C.(1956) "Investigations in sampling of milk records for estimating the total lactation yield of cows."

Unpublished thesis I.C.A.R, New Delhi.

16. Martiny B.(1899;1912;1914) "Different Milking Machine Tests".

Arb dtsch LandwGes. No.37; No.211, No.254.

17. Mc Candlish, A.C. and McVicar, A. (1925) "Are Milk Record Association Results Accurate?"

The Scottish Jour. of Agric., Vol.8, pp.201-205.

18. Mc Carthy, M.D. and Boyle, C. (1946) "The accuracy of monthly, six-weekly and two-monthly weighings in estimating milk yields."

J. Dep. Agric., Eire, 43:5-23.

19. McDowell, J.C. (1927) "Testing cows for production every other month."

Circ. U.S. Dep: Agric. 1.

20. Peterson, W.E. (1925) "Accuracy of one-day and two-day tests".
Jersey Bull.44-(2044, 2058)

21. Gaiz, L.(1927) "Milk production and the 6-5-8 system."
Industr. pecuar. 28 No. 918.

22. Sukhatme, P.V.(1953) "Sampling theory of Survey with Applications."
Indian Soc. of Agri. Res., New Delhi. The Iowa State College Press, Ames Iowa, U.S.A.

23. Tyler, W.J & Chapman, A.B. (1944) "A simplified method of estimating 305-day lactation production."
J. Dairy Sci. 27(6) 463-69.

24. Yapp, W.W.(1919) "A study of the relative reliability of tests for dairy cows."
Bull. Ill. Agri. Exp. Station. 215.

25. Yates, F. (1948) "Systematic Sampling."
Phil. Transac. Roy. Soc., Ser.A, Vol.241, pp.345-78.