Jour. Ind. Soc. Ag. Statistics
44 (3), 1992 : 216–226

Estimation of Lactation Yield from Incomplete Lactation Records

Randhir Singh and P. Narain
IASRI, New Delhi-110 012
(Received : October, 1988)

Summary

For estimating the lactation Yield of cattle and buffaloes daily milk is recorded at systematic time intervals during the entire lactation period for a sample of selected animals. But due to several reasons it may not be always possible to obtain milk records at every time interval for all the selected animals. In the present investigation two alternative estimators of average lactation yield under village conditions have been suggested. The first estimator is based on daily milk records of animals having complete lactation data using appropriate probabilities of obtaining the complete records and the second one makes use of the records of all the selected animals, having complete as well as incomplete lactation records.

Key Words : Lactation yield, time interval, conditional probability.

Introduction

Estimates of average milk yield of cattle and buffaloes maintained in villages are often required for assessing the progress of a development programme aimed at increasing the milk production in the area. Amble *et al* [2] developed a sampling method to estimate the average lactation yield as well as the annual milk production under the village conditions. This procedure requires all the sample milk yield records during the lactation of each animal included in the sample. However, it is not uncommon that some of the records for some of the animals may not be available. This may be due to several reasons, e.g. the selected animals may be sold, transferred or removed due to death during the period of enquiry. Sometimes the enumerator may fail to reach in time to record the milk yield. Thus only a smaller number of animals may ultimately provide the milk records for all the sampled time intervals during the lactation period. The probability of obtaining all the sample milk yield records from different animals depends upon several factors and varies from animal to animal so that the usual estimator based on the sample average of these animals is likely to be biased. Singh and Narain [5] have presented an appropriate estimation procedure based on random sample sizes. In the present study two alternative estimators of average lactation yield have been developed which are unbiased and have smaller variance. The first estimator considers only those animals which provide all the sampled yield records during lactation taking into account the differential probabilities of obtaining such records for the selected animals. The second estimator makes use of the milk records of all the selected animals included in the survey irrespective of whether all the sampled records during lactation are available or not.

2. *Procedure of sampling and estimation*

In order to estimate the average lactation yield, the milk recording is done at systematic intervals throughout the lactation period for a number of selected animals. As the calving of the animals is spread throughout the year it is not possible to obtain the required sample of freshly calved animals at the start of the survey. Therefore, the selection of animals in the sample has to be carried over the average calving interval systematically after up-dating the frame of freshly calved animals at each subsequent time interval. Amble *et al* [2] proposed the plan of selecting a systematic sample of animals which requires the list to be brought up-to-date from time to time before each selection. There-after the milk recording is done at regular intervals of t days throughout the lactation. Such a procedure of systematic sampling both for the animals as well as for the day of milk recording assumes that the selected animal is in milk on the day of milk recording and the sample is evenly distributed over animals calved in different months so that estimates of milk production are not affected by the seasonal variation in calving. But due to several reasons it is not always possible to record milk yield for all the time-intervals for some of the animals. This results in making the estimator biased as well as less efficient. In the present investigation two alternative estimators are proposed to deal with the problem of such incomplete records.

3. *Unbiased estimator based on Complete milk yield records only*

Let y_{ij} denote the milk yield for the j-th time stage of the i-th animal. Let L_i denote the lactation length for i-th animal and t denote the time interval for milk yield recording. The lactation yield of the i-th animal is then given by $Y_i = \Sigma^{L_i} Y_{ij}$. An unbiased estimator of Y_i. is

$$y_{i.} = t \sum_{j=1}^{r_i} y_{ij} = t \ T_{is} \tag{1}$$

where r_i is the number of total sample records and T_{is} is the total of the observed sample milk yield for the i-th animal.

$$V(y_{i.}/i) = t \sum_{s} T_{is}^2 - y_{i.}^2 \ (= v_i \text{ say}) \tag{2}$$

Let p_i denote the probability that the i-th animal selected in the sample would provide the lactation records for all the r_i time-intervals during the lactation period L_i. Let m out of n selected animals provide with records for all the time intervals. Then, following Singh and Narain [5] let the probability of obtaining m animals with records for all the time intervals be

$$p(m) = \Sigma_0 \ \pi_{i}^{m} \ p_i \ \pi_{j \neq i}^{n-m} \ q_j$$

where Σ_0 denotes sum over all $\binom{n}{m}$ combinations and $q_i = 1 - p_i$. The conditional probability of obtaining complete records from the i-th animal given that m animals have complete sample records is given by

$$\pi_i(m) = \frac{p_i \left[\Sigma_1 \ \pi_{i' \neq i}^{m-1} \quad P_{i'} \quad \pi_{j \neq i'}^{n-m} \ q_j \right]}{p(m)}$$

where Σ_1 denotes summation over $\binom{n-1}{m-1}$ combinations excluding the i-th one.

The probability of obtaining complete records from the pair of animals i and j given that m animals have complete sample records is given by

$$\pi_{ij}(m) = \frac{p_i p_j \left[\Sigma_2 \ \pi_{i' \neq i,\, j}^{m-2} \quad P_{i'} \quad \pi_{j' \neq i,\, j,\, i'}^{n-m} \ q_{j}' \right]}{p(m)}$$

where Σ_2 denotes sum over all $\binom{n-2}{m-2}$ combinations excluding i and j. We assume here that the probability of obtaining a sample with no animal having records for all the time intervals i.e. P(O) is very small and negligible so that m is always taken as greater than zero. An unbiased estimate of average lactation yield $\bar{Y}$ is then given by

$$T_1 = \frac{1}{n} \sum_{i=1}^{m} \frac{y_{i.}}{\pi_i(m)} \tag{3}$$

The expected value and the variance of T_1 are given by

$$E(T_1) = E_1E_2E_3(T_1) \text{ and}$$

$$V(T_1) = V_1E_2E_3(T_1) + E_1V_2E_3(T_1) + E_1E_2V_3(T_1)$$

where E_3, E_2, E_1 and V_3, V_2, V_1 denote expected value and variance for given animal i, for fixed m and for all samples of n respectively.

It is seen that T_1 is an unbiased estimator of the lactation yield as

$$E(T_1) = E_1E_2E_3(T_1) = E_1E_2 \sum_{i=1}^{m} \frac{Y_{i.}}{\pi_i(m)} = E_1 \bar{y}_{n.} = \bar{Y}$$

To obtain the variance of T_1 we see that

$$V_1E_2E_3(T_1) = f \Sigma_{s'} y_n^2 - Y_N^2 = V_{sy.} \text{ (say)}$$

where $\Sigma_{s'}$ denotes sum over all such samples and f is the sampling interval for selection of animals. Y_n is total of lactation yield of n selected animals and Y_N is total population milk yield.

$$E_1 V_2E_3(T_1) = \frac{N(n-1)}{n(N-1)} \sum_{i<J}^{N}\sum \left[\pi_i(m)\,\pi_j(m) - \pi_{ij}(m)\right] \left(\frac{Y_{i.}}{\pi_i(m)} - \frac{Y_{j.}}{\pi_j(m)}\right)^2$$

and $$E_1E_2V_3(T_1) = \frac{1}{nM} \sum^{N} \frac{V_i}{\pi_i(m)}$$

so that we may write the variance of T_1, for fixed m, as

$$V(T_1) = V_{sy} + \frac{N(n-1)}{n(N-1)} \sum\sum_{i<j}^{N} \left[\pi_i(m)\ \pi_j(m) - \pi_{ij}(m)\right]$$

$$\left(\frac{Y_{i.}}{\pi_i(m)} - \frac{Y_{j.}}{\pi_j(m)}\right)^2 + \frac{1}{nN} \sum^{N} \frac{V_i}{\pi_i(m)} \tag{4}$$

An approximate estimator of variance of T_1 (by treating the systematic sample of n animals equivalent to a random sample and ignoring the third stage variance i.e. treating the estimated lactation yield of each animal as the true lactation yield) may be written as

$$\hat{V}(T_1) = \frac{1}{n-1}\left[\sum^{m} \frac{y_{i.}^2}{\pi_i(m)} - \frac{1}{n}\left(\sum^{m} \frac{y_{i.}^2}{\pi_i(m)}\right) + \sum\sum_{i<j} \frac{y_{i.}\, y_{j.}}{\pi_i(m)\, \pi_j(m)}\right]^2$$

$$+ \frac{N^2}{n^2} \sum^{m} \sum^{m} \frac{\pi_i(m)\, \pi_j(m) - \pi_{ij}(m)}{\pi_{ij}(m)} \left(\frac{y_{i.}}{\pi_i(m)} - \frac{y_{j.}}{\pi_j(m)}\right)^2$$

The estimator T_1 is based on the assumption that p_i's are known for all the animals. But in general p_i is not known and therefore to make use of the estimator T_1 we have to use some guessed values of P_i's based on the experience and knowledge about the house holder's occupation, breed of the animals, etc. For example some additional information may be obtained regarding whether the house holder intends to sell the animal or wants to keep it for his own use throughout the lactation. Similarly animals of same breed having almost same lactation length may have same P_i.

The estimator T_1 assumes a particularly simple form if we assume all p_i's to be equal (say p) with q = 1–p. In this situation $\pi_i(m)$ reduces to, as expected,

$$\pi_i(m) = \frac{m}{n}$$

The estimator T_1 in this case becomes

$$T_1' = \frac{1}{m} \sum^{m} y_{i.} = \bar{y}_m. \tag{5}$$

and $$V(T_1') = \left[E\left(\frac{1}{m}\right) - \frac{1}{N}\right] S_{y.}^2 + \frac{1}{N}\sum^{N} v_i \; E\left(\frac{1}{m}\right)$$

where

$$E\left(\frac{1}{m}\right) = \left(\frac{1}{np}\right) + \left(\frac{q}{n^2 p^2}\right) \qquad (6)$$

4. *Estimator based on incomplete milk yield records*

The estimator T_1 given by (3) requires the knowledge of p_i for all the animals which is generally not available. The Estimator also does not use the records of (n–m) animals for which milk yield for some time intervals is not available. This results in loss of information. Further as the selection of animals is spread over the calving interval, the survey has to be continued for about three years to obtain complete lactation records of all the animals selected in the sample which requires much time and cost and perhaps will be unusable by the time it is made available.

Therefore we propose an alternative estimator which makes use of all the available information and also can be completed during only one year or so.

Consider a population of N milch animals and let a sample of n animals be selected by simple random sampling without replacement for milk recording at systematic intervals of time t over the lactation period. As the sample is selected out of all the milch animals, some animals may be already in milk and at different lactation stages while some others may calve only after some time. Therefore, all the animals in the sample can not provide record for all the time intervals. Let n' denote the number of animals for which complete lactation records for all the time intervals are available and let n_k' denote the number of additional animals for which the k-th time-interval record is available. Let $n_k = n' + n_k'$ denote the total number of animals for which the k-th time interval records are available.

Now we define

$\bar{y}_.' = \frac{1}{n'} \sum y_{i.}$ the sample mean for y. based on n' animals,

(y. is used for the variable lactation yield) similarly

$\bar{y}_k$, $\bar{y}_{k'}$ = the sample mean for y^k (k-th time interval milk yield) based on n_k and n'_k animals respectively

S_0^2, S_k^2 = the population mean square for y. and y^k

S_{ok}, S_{kl} = the population sum of products for y., y^k and y^k, y^l respectively and

ρ_{ok}, ρ_{kl} = the correlation coefficient between y., y^k and y^k, y^l respectively.

Now as total milk yield y. is related to y^k, the k-th time interval milk yield, an improved estimator of the average lactation yield may be given by

$$T_2 = \bar{y}'_. + \sum^{r} W_k (\bar{y}^k - \bar{y}^{k'})$$

where r is the total number of time intervals at which milk records are available.

or which may be written as

$$T_2 = \bar{y}'. + W' X \qquad (7)$$

W′ is row vector of constants $(W_1, W_2, \ldots, W_r)$ to be chosen such that $V(T_2)$ is minimum and

X is the column vector $(\bar{y}^1 - \bar{y}^{1'}, \ldots, \bar{y}^r - \bar{y}^{r'})'$

The variance of T_2 is given by

$$V(T_2) = V(\bar{y}'_.) + W' \Sigma W + 2 W' Z$$

where Σ is the variance covariance matrix of X, and Z is covariance matrix of $\bar{y}'$. with X.

The variance is minimum when $W = -\Sigma^{-1} Z$.

The minimum variance of T_2 is obtained as

$$V(T_2) = V(\bar{y}'.) - Z' \Sigma^{-1} Z \qquad (9)$$

An estimate of the variance of T_2 can be easily obtained by using the sample values in (9).

To obtain $V(T_2)$ we require $V(\bar{y}'.)$, Σ and Z.

Now

$$V(\bar{y}'.) = V_1E_2(\bar{y}'.) + E_1V_2(\bar{y}'.)$$

$$= \frac{1}{n'N}\sum^{N}\left[\sum_{s} t\ T_{is}^2 - y_{i.}^2\right] + \left(\frac{1}{n'} - \frac{1}{N}\right)S_{y.}^2$$

To obtain Σ and Z we note that

$$V(\bar{y}^k - \bar{y}^{k'}) = S_k^2\left(\frac{1}{n'} - \frac{1}{n_k}\right)$$

$$Cov.(\bar{y}^k - \bar{y}^{k'}, \bar{y}^l - \bar{y}^{l'}) = S_{kl}\left(\frac{n'}{n_kn_l} - \frac{1}{n_k} - \frac{1}{n_l} + \frac{1}{n'}\right)$$

$$= S_{kl}\left(\frac{1}{n'} - \frac{n_k + n_l - n'}{n_k\, n_l}\right)$$

and

$$Cov.(\bar{y}'., \bar{y}^k - \bar{y}^{k'}) = C_1E_2(\bar{y}'., \bar{y}^k - \bar{y}^{k'})$$

$$+ E_1C_2(\bar{y}'., \bar{y}^k - \bar{y}^{k'})$$

$$= C_1E_2(\bar{y}'., \bar{y}^k - \bar{y}^{k'})$$

$$= -S_{ok}\left(\frac{1}{n'} - \frac{1}{n_k}\right)$$

Thus

$$\Sigma = \begin{vmatrix} S_1^2\left(\frac{1}{n'} - \frac{1}{n_2}\right) & S_{12}\left(\frac{1}{n'} - \frac{n_1 + n_2 - n'}{n_1n_2}\right) & \cdots & S_{1r}\left(\frac{1}{n'} - \frac{n_1 + n_r - n'}{n_1n_r}\right) \\ & S_2^2\left(\frac{1}{n'} - \frac{1}{n_2}\right) & \cdots & S_{2r}\left(\frac{1}{n'} - \frac{n_2 + n_r - n'}{n_2n_r}\right) \\ & & \cdots & \\ & & \cdots & \\ & & \cdots & S_r^2\left(\frac{1}{n'} - \frac{1}{n_r}\right) \end{vmatrix}$$

and

$$Z = -\left[S_{01}\left(\frac{1}{n'} - \frac{1}{n_1}\right), S_{02}\left(\frac{1}{n'} - \frac{1}{n_2}\right), \ldots, S_{0r}\left(\frac{1}{n'} - \frac{1}{n_r}\right)\right]$$

From (9) it can be seen that the estimator T_2 is always more efficient as compared to the estimator based only on the sample mean of data for animals having lactation records for all the time-intervals. The reduction in variance due to the use of estimator T_2 instead of $\bar{y}'.$ as obtained from (9) is given by

$$V(\bar{y}'.) - V(T_2) = Z' \Sigma^{-1} Z \tag{10}$$

5. *Empirical investigation*

In order to study the relative efficiency of the estimator T_2 which makes use of all the available data, consider an empirical inveistation. Data for this study has been taken from a survey entitled 'A Study of Impact of Milk Supply Schemes on Rural Economy in milk collection areas of Madhavaram Milk Supply Schemes, Madras'. conducted by the Indian Agricultural Statistics Research Institute, ICAR during the years 1979-80.

For the purpose of present study milk yield records obtained at monthly intervals over a period of one year for 44 cattle selected randomly from two villages have been used. Out of the total of these 44 cattle, data for all time stages during the entire lactation period could be obtained for only 11 cattle. The complete data is given in the Annexure.

Now, if data of only such cattle is used whose lactation records are complete then, we can use the records of only 11 cattle and the estimator of average lactation yield is provided by $\bar{y}'.$, the sample mean for total lactation yield of these cattle. The estimate of variance of $\bar{y}.$ is 688.50. Using the estimator T_2 proposed in (7) to make use of the entire available data, the records of all the 44 animals are considered. The estimate of variance of T_2 is 179.57 which shows that the estimated variance reduces to almost one fifth by use of estimator T_2 which a reasonable indicator that T_2 is expected to be quite efficient as compared to $\bar{y}'.$. Assuming the probability of obtaining complete lactation records from each cattle equal, the estimator T_1' given by (5) reduces to estimator $\bar{y}'.$, therefore for the empirical study the estimator $\bar{y}'.$ has been used.

ANNEXURE

Daily Milk Yield records at monthly intervals during the Laction (100 Gms.)

Sl. No.	Lactation stage											
	1	2	3	4	5	6	7	8	9	19	11	12
1.	18	18	20	20	20	20	20	20	15	10	08	05
2.	10	10	10	10	10	10	10	10	..	..	..	..
3.	..	15	15	10	10	10	10	10	10	10	05	05
4.	15	20	20	20	20	20	20	20	..	..	..	..
5.	10	10	10	10	10	10	10	..	..	..	..	..
6.	..	15	15	15	15	15	15	..	10	10	10	10
7.	15	15	10	10	10	10	10	10	10	10	10	05
8.	15	15	10	10	10	10	10	10	10	10	10	10
9.	..	25	22	22	20	20	20	20	20	10	10	10
10.	..	..	..	..	..	..	..	15	15	15	10	05
11.	..	..	..	..	..	..	15	15	15	10	05	..
12.	10	10	15	15	..	..	..	..	..	..	..	..
13.	..	..	..	15	14	14	15	10	10	10	10	08
14.	..	..	..	..	25	20	20	20	20	10	10	10
15.	..	..	..	..	20	22	22	20	20	20	20	20
16.	..	..	..	20	24	24	20	20	15	15	10	10
17.	..	..	..	..	..	..	15	14	20	10	10	..
18.	15	15	10	10	10	10	10	10	10	10	05	05
19.	..	..	..	..	..	..	..	20	20	10	05	10
20.	..	..	..	..	..	..	..	..	15	20	10	05
21.	10	15	15	15	15	..	..	..	..	..	..	..
22.	10	10	10	15	15	10	10	10	10	08	06	06
23.	15	15	15	10	10	10	..	10	10	10	..	..
24.	10	15	15	15	15	15	..	..	..	..	..	..
25.	25	25	25	25	25	20	20	20	20	20	20	13
26.	..	..	..	20	20	20	20	20	20	20	10	06
27.	..	..	..	..	..	10	10	10	10	10	10	10
28.	..	..	10	10	10	10	10	10	10	07	05	..
29.	..	..	..	..	20	20	20	20	20	20	20	10
30.	10	10	10	10	10	10	10	10	10	10	10	05
31.	..	..	..	15	15	15	15	25	15	15	10	12
32.	..	..	..	25	25	25	25	..	25	25	20	16
33.	..	..	..	..	40	40	40	40	40	30	30	25
34.	..	..	..	40	40	40	40	40	40	40	30	10
35.	40	40	40	45	45	45	40	40	35	28	20	10
36.	40	40	40	40	40	40	35	30	30	25	10	10
37.	..	..	..	30	30	..	30	30	30	30	..	20
38.	20	20	20	20	20	20	20	20	20	13	06	05
39.	..	..	..	25	..	45	40	40	40	40	..	25
40.	..	..	..	30	30	30	30	30	30	20	25	10
41.	20	25	25	25	25	25	..	..	..	..	..	..
42.	..	..	..	..	30	30	30	30	30	30	30	20
43.	..	..	..	..	30	30	30	30	25	25	25	15
44.	15	10	10	10	10	10	10	10	10	08	06	15

REFERENCES

[1] Amble, V.N. and Rajagopalan, M, 1960. Further studies in sampling daily milk yields. *Indian Jr. Vet. Sci.* **30(4)**, 272-86.

[2] Amble, V.N. and Rajagopalan, M and Gill, B.S., 1959. Sampling of daily milk records for estimating lactation yield of cows'. *Indian Jr. Vet. Sci.*, **29(4)** 13-26.

[3] Aneja, K.G. and Marutiram, B., 1974. 'A sampling plan for the measurement of changes in level of milk production in area under development' *Ind. Jour. of Anim. Sci.* **44** No.2, 70-75.

[4] A study of Impact of milk supply schemes on Rural Economy in milk collection areas of Madhavaram milk supply schemes, Madras (1979-80). Project Report, IASRI publication.

[5] Singh Randhir and Narain P., 1989. Method of estimation from samples with random sample sizes. *J.S.P.I.* **23**, 217-225.