COMPARISON OF SOME ESTIMATORS OF ANNUAL MILK PRODUCTION

BY D. SINGH V. V. R. MURTY AND B. B. P. S. GOEL
Institute of Agricultural Research Statistics, New Delhi
(Received in February, 1968)

1. INTRODUCTION

The available official estimates of annual production of milk in the country are those worked out by the Directorate of Marketing & Inspection on the basis of the market surveys conducted by them (Chowdhary and Narang; 1962). These estimates are of limited utility since the data used were not those collected through objective enquiries (Singh and Murty, 1963).

During the Second Plan period, the Institute of Agricultural Reserch Statistics had conducted a series of pilot sample surveys in typical areas of the different animal husbandry regions of the country with the objects of evolving a suitable sampling procedure for securing precise estimates of milk production and study of the practices of feeding and management of cattle and buffaloes in the regions (Panse, Singh and Murty, 1964, 1966). Such investigations were conducted in Punjab, eastern districts of Uttar Pradesh and Gujarat and costal districts and adjoining areas of Andhra Pradesh and coastal districts of Orissa (Amble, Murty, Sathe and Goel, 1964). During the Third Plan period, the first three investigations were repeated after a lapse of 5 years for the changes in the level of milk production rnd in the bovine practices that might have taken place during the intervening period.

In each of the surveys, data were collected over a whole year covering each of the three seasons, *viz.*, summer, rainy and winter, using appropriate sampling methods. On every working day of the survey, a representative sample of cows and buffaloes which were in milk on that day was selected and the milk yield of these animals was recorded by physical weighment on the spot. Such data collected by appropriate sampling procedures over the whole year permitted to work out objective estimates of annual milk production in the tract.

The annual milk production was thus obtained as—estimate of average milk yield per day per animal in milk multiplied by the estimate of the average number of animals in milk in the year and the number of days in the year (365). Estimates for different seasons were also obtained by using a similar procedure.

2. Scope of Study

For a particular type of sampling design used in a survey, a number of estimates of annual milk production can be formulated. From this wide class of estimates, the most suitable estimate could be chosen after taking into consideration the relative efficiency of the estimate besides simplicity in its calculation. On certain occasions, a biased estimate may have to be preferred to an estimate which is unbiased provided the bias is small and can be estimated from the sample itself. Keeping these considerations in view, several estimates of number of animals in milk and annual milk production have been studied and their relative efficiency has been examined. This paper gives the results of such a study. In working out the estimates of the number of animals in milk and annual milk production, ratio method of estimation has been used. These estimates, although biased, are consistent. An estimate of the relative bias has also been worked out for each of the estimates and the results are given in the paper. The assumption made in arriving at the various expressions in the estimates is that the sample size is large and, therefore, only first approximation has been used in arriving at them. Thus the contribution by second and higher order terms has been neglected. The assumption of large sample is possibly valid, the sampling fraction at the first stage varying from 20 to 25 per cent in different strata. The correlation between the character studied and the auxiliary character used is assumed and seen to be high. For obtaining the estimates of variance of number of animals in milk or milk production sample estimates were substituted for the corresponding population values. Similarly, a preliminary study revealed that the contribution of the bias at the second stage in the relative bias of number of animals in milk or milk production was negligible as compared to the first stage component and the second stage component of bias has, therefore, been ignored.

The various mathematical expressions of the estimates, their relative bias and variance are indicated in Section 4. A numerical illustration has also been presented making use of the data relating to rural area collected from the surveys of Punjab and Uttar Pradesh. The sampling design adopted in these states is indicated in Section 3.

3. Sampling Design of Pilot Sample Surveys in Punjab and Uttar Pradesh

(*i*) *Sampling Design* : The sampling design adopted in each of the states was one of stratified multistage random sampling. The tract was divided into three or four zones and the zones were adopted as strata. The basis of stratification was geographical contiguity, climatic conditions, animal husbandry practices, etc. The unit at the primary stage of sampling (PSU) was a tehsil. A cluster of three neighbouring villages/town, a cluster of three adjacent households and an animal in milk were the units at the second, third and ultimate stages of sampling. From each stratum, a sample of five tehsils was selected with equal probability and without replacement and these tehsils were followed for observation throughout the 12 months' period of the survey. From each tehsil, a sample of two clusters of villages or one such cluster and two towns were selected in each season. Selection of a cluster was done by first selecting a village with equal probability and without replacement from the total population of villages in the tehsil and attaching to this village, two more villages selected at random out of those which were within a radius of three kilometers. Selection of the sample at the subsequent stages was done with equal probability and without replacement. In case of urban area, a further stage was introduced by first selecting a town from the tehsil and then selecting a ward from the selected town. Thus, from each tehsil, two clusters of three villages each or one such cluster and two towns with one ward in each selected town were observed for field work and this sample was observed in each month of the season. During each month from each selected village, a sample of 2 to 4 clusters of 3 households each was selected for recording detailed data. These clusters were those formed by first selecting a household having animals and clubbing it with two similar adjoining households. From each selected household two animals in milk were selected for recording milk yield, while all the animals in the household were observed for recording other detailed data.

(*ii*) *Programme of field work* : An enumerator was located in each selected tehsil and he was required to collect the primary data from the sample of villages/towns selected from the tehsil. He would visit each of the two clusters of villages or one such cluster and two towns in his jurisdiction every month, spending a fortnight in each cluster. In the course of the first month, the enumerator would devote the first 8 days of the fortnight for making a complete enumeration of the households in the selected villages and collect the necessary basic data for preparing the sampling frame. In the remaining part of the fortnight, he would select two clusters of three households from each village and record detailed data in respect of each selected household at the rate of one cluster per day. In case of a ward, the procedure was exactly the same except in that he would stay only for a week in each ward and spend 4 days for a complete enumeration and subsequent two days in the week for collection of detailed data. During each of the subsequent months, four clusters were selected in each village for collection of detailed data, while in case of a ward, five such clusters were selected. Besides these clusters, the enumerator would also select 8 more clusters of three households each in case of a village and 10 clusters in case of a ward for recording changes in the number of bovine stock in the period intervening the first and the subsequent visits.

(*iii*) *Data collected* : The detailed data collected from each of the selected households included information on milk yield for each of the milkings in the day of two animals in milk selected at random, out of all the animals in the household. Data were also recorded on quantity and composition of feeds supplied to the animals and particulars such as breed, the order and stage of lactation of animals in milk, prices of purchase and sale of animals, veterinary aid availed of, utilisation of milk, procurement of feeds, utilisation of dung, etc.

4. Estimates of Number of Animals in Milk and Total Milk Production

It has been mentioned in the previous section that from each selected tehsil, a sample of two clusters each consisting of three adjacent villages was selected. Utilising the mean squares between clusters within tehsils and between villages within clusters in the analysis of variance, an estimate of the intraclass correlation coefficient between the village means was worked out and this was found to be very small, *viz.*, 0 1, for each of the surveys. Thus for further study, the procedures used were the same as for a random sample of villages from all the villages of the selected tehsil.

4·1. Notation :

Let y_{ijrklm} be the milk yield of 'm'th animal, in 'l'th household, 'k'th village of 'i'th tehsil during 'r'th round of 'j'th season.

Further notation may be defind as follows :—

	Population	*Sample*
Number of animals in milk in 'l'th household, 'k'th village of 'i'th tehsil during 'r'th round of 'j'th season.	M_{ijrkl}	m_{ijrkl}
Number of households in 'k'th village of 'i'th tehsil during 'r'th round of 'j'th season.	H_{ijrk} (H_{ik} for all 'j' and 'r')	$hijrk$
Number of villages in 'i'th tehsil during 'j'th season.	V_{ij} (or simply V_i)	v_{ij}
Number of rounds in 'j'th season.	d_j (d_j=3, 4 or 5)	d_j
Relative length of the 'j'th season.	$p_j = \frac{d_j}{12}$ $\quad \sum_{j=1}^{3} p_j = 1$	
Number of tehsils in the stratum.	T	t
Number of seasons.	3	3

Further, let. $$M_{ijrk} = \sum_{l=1}^{H_{ik}} M_{ijrkl}$$

$$M_{ijk} = \frac{\sum_{r=1}^{d_j} M_{ijrk}}{d_j}$$

$$M_{ij} = \sum_{k=1}^{V_i} M_{ijk}$$

$$\overline{M}_{ij} = \frac{M_{ij}}{V_i}$$

$$M_j = \sum_{i=1}^{T} M_{ij}$$

$$M_i = \sum_{j=1}^{3} p_j M_{ij}$$

$$M = \sum_{i=1}^{T} M_i$$

$$\overline{M} = \frac{M}{T}$$

It may be noted that the 'M's are true values of the number of animals in milk but these values, with the exception of 'M_{ijk}' namely the number of animals in milk in a sampled village enumerated during a season, are not known. However, similar data on the number of animals in milk and milch animals according to a quinquennial Livestock Census are known for each village and tehsil in the population. The notation used to denote the number of animals in milk or milch animals according to livestock census is given below.

Let M'_{ik} be the census figure of number of animals in milk in 'k'th village of 'i'th tehsil and M''_{ik} the corresponding figure for number of milch animals and

$$M'_i = \sum_{K=1}^{V_i} M_{ik}, \qquad M_i'' = \sum_{k=1}^{V_i} M''_{ik}$$

$$\overline{M_i'} = \frac{M_i'}{V_i} \qquad \overline{M_i''} = \frac{M_i''}{V_i}$$

$$M' = \sum_{i=1}^{T} M_i' \qquad M'' = \sum_{i=1}^{T} M_i''$$

$$\overline{M'} = \frac{M'}{T} \qquad \overline{M''} = \frac{M''}{T}$$

4·2. ESTIMATE OF THE NUMBER OF ANIMALS IN MILK

Three different estimates of M, denoted by $\hat{M}_{(1)}$, $\hat{M}_{(2)}$ and $\hat{M}_{(3)}$ were worked out and the estimates are given below.

$$(i) \quad \hat{M}_{(1)} = M' \times \frac{\sum_i^t \hat{M}_{i(1)}}{\sum_i^t M_i'}$$

where

$$\hat{M}_{i(1)} = \sum_{j=1}^{3} p_j \hat{M}_{ij(1)}$$

and

$$\hat{M}_{ij(1)} = M'_i \times \frac{\sum_k^{v_{ij}} M_{ijk}}{\sum_k^{v_{ij}} M'_{ik}}$$

$$(ii) \quad \hat{M}_{(2)} = M' \times \frac{\sum_i^t \hat{M}_{i(2)}}{\sum_i^t M'_i}$$

where

$$\hat{M}_{i(2)} = \sum_{j=1}^{3} p_j \hat{M}_{ij(2)}$$

and

$$\hat{M}_{ij(2)}=\frac{V_i}{v_{ij}}\sum_{k}^{v_{ij}} M_{ijk}$$

$$(iii)\quad \hat{M}_{(3)}=H\times\frac{\sum_{i}^{t}\hat{M}_{i(3)}}{\sum_{i}^{t} H_i}$$

where 'H_i' is the number of households in 'i'th tehsil according to Human population census, and

$$H=\sum_{i=1}^{T} H_i$$

$$\hat{M}_{i(3)}=\sum_{j=1}^{3} p_j\,\hat{M}_{ij(3)},$$

$$\hat{M}_{ij(3)}=H_i\times\frac{\sum_{k}^{v_{ij}} M_{ijk}}{\sum_{k}^{v_{ij}} H_{ik}}$$

The estimates $\hat{M}_{(1)}$, $\hat{M}_{(2)}$ and $\hat{M}_{(3)}$ being ratio estimates are biased, but consistent.

Relative bias in $\hat{M}_{(1)}$ is given by

$$q\times\left\{\frac{\sum_{i=1}^{T}(M_i'-\overline{M}')^2}{\overline{M}'^2}-\frac{\sum_{i=1}^{T}(M_i-\overline{M})(M_i'-\overline{M}')}{\overline{M}\,\overline{M}'}-\frac{\sum_{i=1}^{T}(M_i'-\overline{M}')(B_i-\overline{B})}{\overline{M}\,\overline{M}'}\right\}+\frac{\overline{B}}{\overline{M}}\qquad\ldots(1)$$

where $q=\frac{T-t}{Tt}\cdot\frac{1}{(T-1)}$, $\overline{B}=\frac{1}{T}\sum_{i=1}^{T} B_i$ and B_i is the relative bias in $\hat{M}_{i(1)}$. Neglecting bias at the second stage, estimate of $R.B.$ $(\hat{M}_{(1)})$ can be shown to be approximately equal to

$$q=\left\{\frac{\sum_{i}^{t}(M_i'-\hat{\overline{M}}')^2}{\hat{\overline{M}}'^2}-\frac{\sum_{i}^{t}(\hat{M}_{i(1)}-\hat{\overline{M}})(M_i'-\hat{\overline{M}}')}{\hat{\overline{M}}\,\hat{\overline{M}}'}\right\}\qquad\ldots(2)$$

where

$$q' = \frac{T-t}{Tt}\,\frac{1}{t-1} \text{ and } \hat{\overline{M'}} = \sum_{i}^{t} \frac{M_i'}{t}$$

$$V(\hat{M}_{(1)}) = T^2 q \sum_{i=1}^{T} \left(M_i - \frac{\overline{M}}{\overline{M'}} M'_i\right)^2 + \frac{T}{t} \sum_{i=1}^{T} V^2_i \sum_{j=1}^{3} \frac{p^2_j}{v_{ij}(V_i-1)} \sum_{k=1}^{V_i} \left(M_{ijk} - \frac{\overline{M}_{ij}}{\overline{M'}_i} M'^2_{ijk}\right)^2 \quad \ldots(3)$$

$$\text{Est. } V(\hat{M}_{(1)}) = T^2 q' \sum_{i}^{t} \left(\hat{M}_{i(1)} - \frac{\hat{\overline{M}}}{\hat{\overline{M'}}} M_i'\right)^2 + \frac{T}{t} \sum_{i}^{t} V^2_i \sum_{j=1}^{3} \frac{p_j^2}{v_{ij}(v_{ij}-1)} \sum_{k=1}^{v_{ij}} \left(M_{ijk} - \frac{\hat{\overline{M}}_{ij(1)}}{\overline{M}_i'} M_{ijk}'\right)^2 \quad \ldots(4)$$

Relative bias in $\hat{M}_{(2)}$ is given by

$$q \times \left\{ \sum_{i=1}^{T} \frac{(M_i' - \overline{M'})^2}{\overline{M'}^2} - \sum_{i=1}^{T} \frac{(M_i' - \overline{M'})(M_i - \overline{M})}{\overline{M'}\,\overline{M}} \right\} \quad \ldots(5)$$

Estimate of Relative bias in $\hat{M}_{(2)}$ is approximately equal to

$$q \times \left\{ \sum_{i}^{t} \frac{(M_i' - \hat{\overline{M'}})^2}{\hat{\overline{M'}}^2} - \sum_{i}^{t} \frac{(M_i' - \hat{\overline{M'}})(\hat{\overline{M}}_{i(2)} - \hat{\overline{M}}_{(2)})}{\hat{\overline{M}}\,\hat{\overline{M'}}_{(2)}} \right\} \quad \ldots(6)$$

$$V(\hat{M}_{(2)}) = T^2 q \sum_{i=1}^{T} \left\{ M_i - M_i' \frac{\overline{M}}{\overline{M'}} \right\}^2 + \frac{T}{t} \sum_{i=1}^{T} V_i^2 \sum_{j=1}^{3} \frac{p_j^2}{v_{ij}(V_i-1)} \sum_{k=1}^{V_i} (M_{ijk} - \overline{M}_{ij})^2 \ldots(7)$$

$$\text{Est. } V(\hat{M}_{(2)}) = T^2 q \sum_{i}^{t} \left\{ \hat{M}_{i(2)} - M_i' \frac{\hat{\overline{M}}_{(2)}}{\hat{\overline{M'}}} \right\}^2 + \frac{T}{t} \sum_{i}^{t} V_i^2 \sum_{j=1}^{3} \frac{p_j^2}{v_{ij}(v_{ij}-1)} \sum_{k}^{v_{ij}} (M_{ijk} - \hat{\overline{M}}_{ij})^2 \quad \ldots(8)$$

The relative bias in $\hat{M}_{(3)}$ and $V(\hat{M}_{(3)})$ and their estimates will be given by expressions quite similar to those of $\hat{M}_{(1)}$ except in that M_{ik}', M_i', $\overline{M}_i'$, etc., are to be substituted by H_{ik}, H_i, $\overline{H}_i$, etc., res-

...tes $\hat{M}_{(1)}$, $\hat{M}_{(2)}$ and $\hat{M}_{(3)}$ together with the estimates ...bias and variances worked out from the data of the ... and Eastern Uttar Pradesh are given in Table 1.

TABLE 1

Estimate of number of animals in milk, its relative bias and percentage standard error

State	Category	Stratum	$\hat{M}_{(1)}$			$\hat{M}_{(2)}$			$\hat{M}_{(3)}$		
			No. of animals (000)	R.B. (per cent)	S.E. (per cent)	No. of animals (000)	R.B. (per cent)	S.E. (per cent)	No. of animals (000)	R.B. (per cent)	S.E. (per cent)
Punjab	Cows	I	211	0·03	4·7	211	1·40	16·8	248	-0·10	7·0
		II	197	-0·66	4·2	170	1·10	17·0	173	-0·20	6·4
		III	231	0·02	1·1	168	1·00	16·6	176	0·40	7·5
		Overall	639	-0·19	2·1	549	1·20	9·8	597	0·02	4·1
	Buffaloes	I	387	0·20	2·5	471	0·50	9·1	362	-0·50	20·8
		II	305	0·02	2·8	228	-0·80	13·2	275	0·50	3·9
		III	386	-0·40	4·8	342	1·00	19·9	401	-1·10	12·7
		Overall	1078	-0·10	2·1	1041	0·40	8·2	1038	-0·50	8·8
[illegible]r Pradesh	Cows	I	268	0·70	13·6	334	2·03	17·0	338	0·25	15·9
		II	215	-0·09	17·5	227	-2·78	20·8	196	2·28	57·6
		III	252	-0·08	8·7	230	-2·01	11·8	223	0·17	14·1
		IV	191	0·35	3·9	171	-0·91	7·5	205	-0·09	7·9
		Overall	926	0·23	6·2	962	-0·59	8·3	962	0·57	13·5
	[illegible]	I	263	1·54	14·1	265	2·14	17·9	229	0·24	19·8
		II	208	0·72	10·7	194	1·38	20·0	192	1·21	15·7
		III	213	0·13	6·2	209	1·14	16·3	290	-0·56	8·5
		IV	229	0·08	2·6	207	-0·49	16·3	260	0·25	8·7
		Overall	913	0·66	5·0	875	1·10	8·9	971	0·19	6·6

It is seen from the above that $\hat{M}_{(1)}$ is the estimate with minimum relative bias and standard error. $\hat{M}_{(1)}$ has, therefore, been preferred to $\hat{M}_{(2)}$ and $\hat{M}_{(3)}$ for estimating the number of animals in milk in the year.

4·3. Estimate of Milk Production per Day

Let $P=M\bar{y}$ be the total milk production per day in a stratum. The following estimates of P are considered.

1. $$\hat{P}_{(1)}=\hat{M}_{(1)}\times\frac{\sum_{i}^{t} M_i'' \bar{y}_i}{\sum_{i}^{t} M_i''}$$

2. $$\hat{P}_{(2)}=\hat{M}_{(1)}\times\frac{\sum_{i}^{t} \hat{M}_i \bar{y}_i}{\sum_{i}^{t} \hat{M}_i}$$

$$\left[\text{where } \hat{M}_i=\sum_{j=1}^{3} p_j \hat{M}_{ij} \text{ and } \hat{M}_{ij}=M_i'\times\frac{\sum_{k}^{v_{ij}} M_{ijk}}{\sum_{k}^{v_{ij}} M'_{ik}}\right]$$

3. $$\hat{P}_{(3)}=\hat{M}_{(1)}\times\frac{1}{t}\sum_{i}^{t} \bar{y}_i$$

4. $$\hat{P}_{(4)}=\hat{M}_{(1)}\times\frac{\sum_{i}^{t} M_i' \bar{y}_i}{\sum_{i}^{t} M_i'}$$

where

$$\bar{y}_i=\frac{\sum_{j=1}^{3} p_j \hat{M}_{ij} \bar{y}_{ij}}{\sum_{j=1}^{3} p_j \hat{M}_{ij}}$$

$$\bar{y}_{ij}=\frac{\sum_{k}^{v_{ij}} M_{ijk} \bar{y}_{ijk}}{\sum_{k}^{v_{ij}} M_{ijk}}$$

and $\bar{y}_{ijk}$ is the average milk yield per day per animal in milk in 'k'th village of 'i'th tehsil during 'j'th season.

Relative bias of $\hat{P}_{(1)}$ can be shown to be approximately equal to

$$q\times\left\{\frac{\sum_{i=1}^{T}(M_i+B_i-\overline{M}-\overline{B})\{M_i''(\overline{Y}_i+\lambda_i)-\frac{1}{T}\sum_{i=1}^{T}M_i''(\overline{Y}_i+\lambda_i)\}}{\overline{M}\,\overline{M}''\,\overline{Y}}-\frac{\sum_{i=1}^{T}(M_i+B_i-\overline{M}-\overline{B})(M_i'-\overline{M}')}{\overline{M}\,\overline{M}'}\right.$$

$$-\frac{\sum_{i=1}^{T}\{M_i''(\overline{Y}_i+\lambda_i)-\frac{1}{T}\sum_{i=1}^{T}M_i''(\overline{Y}_i+\lambda_i)\}(M_i''-\overline{M}'')}{\overline{M}''^2\,\overline{Y}}$$

$$-\frac{\sum_{i=1}^{T}(M_i+B_i-\overline{M}-\overline{B})(M_i''-\overline{M}'')}{\overline{M}\,\overline{M}''}+\frac{\sum_{i=1}^{T}(M_i'-\overline{M}')(M_i''-\overline{M}'')}{\overline{M}'\,\overline{M}''}$$

$$-\frac{\sum_{i=1}^{T}\{M_i''(\overline{Y}_i+\lambda_i)-\frac{1}{T}\sum_{i=1}^{T}M_i''(\overline{Y}_i+\lambda_i)\}(M_i'-\overline{M}')}{\overline{M}''\,\overline{Y}\,\overline{M}'}$$

$$\left.+\frac{\sum_{i=1}^{T}(M_i'-\overline{M}')^2}{\overline{M}'^2}+\frac{\sum_{i=1}^{T}(M_i''-\overline{M}'')^2}{\overline{M}''^2}\right\}+\frac{\overline{B}}{\overline{M}}$$

$$+\frac{\overline{B}\left\{\frac{1}{T}\sum_{i=1}^{T}M_i''(\overline{Y}_i+\lambda_i)-\overline{M}''\,\overline{Y}\right\}}{\overline{M}\,\overline{M}''\,\overline{Y}}+\frac{\frac{1}{T}\sum_{i=1}^{T}M_i''(\overline{Y}_i+\lambda_i)-\overline{M}''\,\overline{Y}}{\overline{M}''\,\overline{Y}} \quad \ldots(9)$$

Relative mean square error of $\hat{P}_{(1)}$ can be shown to be approximately equal to

$$q\times\left\{\frac{\sum_{i=1}^{T}(M_i+B_i-\overline{M}-\overline{B})^2}{\overline{M}^2}+\frac{\sum_{i=1}^{T}(M_i'-\overline{M}')^2}{\overline{M}'^2}\right.$$

$$+\frac{\sum_{i=1}^{T}\{M_i''(\overline{Y}_i+\lambda_i)-\frac{1}{T}\sum_{i=1}^{T}M_i''(\overline{Y}_i+\lambda_i)\}^2}{(\overline{M}''\,\overline{Y})^2}$$

$$-2\frac{\sum_{i=1}^{T}(M_i+B_i-\overline{M}-\overline{B})(\overline{M}_i'-\overline{M}')}{\overline{M}\,\overline{M}'}$$

$$+\sum_{i=1}^{T}\frac{(M_i''-\overline{M}'')^2}{\overline{M}''^2}$$

$$-2\frac{\sum_{i=1}^{T}(M_i''-\overline{M}'')(M_i+B_i-\overline{M}-\overline{B})}{\overline{M}''\overline{M}}$$

$$+2\frac{\sum_{i=1}^{T}\{M_i'+B_i-\overline{M}-\overline{B}\}\{M''_i(\overline{Y}_i+\lambda_i)-\frac{1}{T}\sum_{i=1}^{T}M_i''(\overline{Y}_i+\lambda_i)\}}{\overline{M}\,\overline{M}''\,\overline{Y}}$$

$$-2\frac{\sum_{i=1}^{T}(M_i'-\overline{M}')\{M_i''(\bar{y}_i+\lambda_i)-\frac{1}{T}\sum_{i=1}^{T}M_i''(\overline{Y}_i+\lambda_i)\}}{\overline{M}'\,\overline{M}''\,\overline{Y}}$$

$$+2\frac{\sum_{i=1}^{T}(M_i'-\overline{M}')(M_i''-\overline{M}'')}{\overline{M}'\,\overline{M}''}$$

$$-2\frac{\sum_{i=1}^{T}(M_i''-\overline{M}'')\{M_i''(\overline{Y}_i+\lambda_i)-\frac{1}{T}\sum_{i=1}^{T}M_i''(\overline{Y}_i+\lambda_i)\}]}{\overline{M}''^2\,\overline{Y}}\Bigg\}$$

$$+\frac{1}{Tt}\left[\frac{\sum_{i=1}^{T}V(\hat{M}_i)+\overline{B}^2}{\overline{M}^2}\right.$$

$$\left.+\frac{\sum_{i=1}^{T}M_i''^2V(\bar{y}_i)+\left\{\frac{1}{T}\sum_{i=1}^{T}M_i''(\overline{Y}_i+\lambda_i)-\overline{M}''\overline{Y}\right\}^2}{(\overline{M}''\overline{Y})^2}\right]$$

$$+2\overline{B}\frac{\left[\frac{1}{T}\sum_{i=1}^{T}M_i''(\overline{Y}_i+\lambda_i)-\overline{M}''\overline{Y}\right]}{\overline{M}\,\overline{M}''\,\overline{Y}} \qquad \text{...(10)}$$

where B_i and λ_i are the biases in the estimates of M_i and $\overline{Y}_i$ which are ratio estimates. As has already been mentioned in section 2, the biases at the second stage, *viz.*, B_i and λ_i have been ignored in further discussion.

Further a study of the correlation coefficients between different estimates of numbers ($\hat{M}$, M', M''), and $\bar{y}$ and between different estimates of numbers and production ($\hat{M}\bar{y}$, $M'\bar{y}$ and $M''\bar{y}$) and the coefficients of variation was made. The estimated values of the correlation coefficients and coefficients of variation obtained for the Punjab survey are given in Table 2. The correlation coefficients and the coefficients of variation have been culculated ignoring the strata. As mentioned earlier, the intraclass correlation between villages within clusters was small. Thus the procedure adopted for working out the correlation coefficients or coefficients of variation were the same as those usually adopted for a random sample.

It may be seen from Table 2 that the correlation coefficients as also their *C.V*'s, between the first three pairs *viz.*, $\hat{M}, M'$; $\hat{M}, M''$ and M', M'' are of the same order and may be denoted by ρ_1 and C_1, Similarly, the values of the fourth and fifth pairs, *viz*, $\hat{M}$, $\bar{y}$ and M', $\bar{y}$, are of the same order and the correlation coefficients and *C.V's* may be denoted by ρ_2 and C_2. Similarly, the correlation coefficients between the remaining pairs of characters, *viz.*, $\hat{M}$, $\hat{M}''\bar{y}$; M', $M''\bar{y}$; M'', $M''\bar{y}$; M', $M'\bar{y}$; M', $\hat{M}\bar{y}$ and M'', $M\bar{y}$ and also their *C.V's* are of a different type and are of the same order. The latter group of coefficients of correlation and coefficients of variation may be denoted by ρ_3 and C_3 respectively. The relative bias and relative mean square error of $\hat{P}_{(1)}$ will then reduce to the expressions given below :—

$$R.B.(\hat{P}_{(1)}) = \frac{T-t}{Tt}\left[C_1^2(2-\rho_1)-\rho_3C_1C_3\right] + \frac{\frac{1}{T}\sum_{i=1}^{T} M_i''\overline{Y}_i - \overline{M}''\overline{Y}}{\overline{M}''\overline{Y}} \qquad \ldots(11)$$

$$R.M.S.E.(\hat{P}_{(1)}) = \frac{T-t}{Tt}\left[C_1^2(3-2\rho_1)+C_3(C_3-2\rho_3C_1)\right] + \frac{1}{Tt}\sum_{i=1}^{T}\frac{V(\hat{M}_i)}{\overline{M}^2} + \frac{1}{Tt}\frac{\sum_{i=1}^{T} M_i''V(\bar{y}_i)}{(\overline{M}''\overline{Y})^2} + \frac{\left(\frac{1}{T}\sum_{i=1}^{T} M_i''\overline{Y}_i - \overline{M''}\overline{Y}\right)^2}{(\overline{M}''\overline{Y})^2} \qquad \ldots(12)$$

Under the assumptions made above, the relative bias and relative mean square error of the other three estimates can be shown to be equal to

TABLE 2

Estimates of coefficient of variation and coefficient of correlation

Characters	COWS			BUFFALOES		
	Cofficient of variation (1)	*Coefficient of variation* (2)	*Correlatior coefficient*	*Coefficient of variation* (1)	*Coefficient of variation* (2)	*Correlation coefficient*
1. $\hat{M}, M'$	44·5	40·4	0·98	45·0	42·8	0·97
2. $\hat{M}, M''$	44·5	36·6	0·91	45·0	38·0	0·93
3. M', M''	40·4	36·6	0·96	42·8	38·4	0·96
4. $\hat{M}, \bar{y}$	44·5	18·3	0·27	45·0	21·2	0·66
5. $M', \bar{y}$	40·4	18·3	0·18	42·8	21·2	0·53
6. $\hat{M}, M'' \bar{y}$	44·5	43·2	0·93	45·0	50·0	0·96
7. $M', M'' \bar{y}$	40·4	43·2	0·92	42·8	50·0	0·92
8. $M'', M'' \bar{y}$	36·6	43·2	0·88	38·4	50·0	0·91
9. $\hat{M}, M' \bar{y}$	44·5	48·8	0·95	45·0	54·9	0·97
10. $M', M' \bar{y}$	40·4	48·8	0·92	42·8	54·9	0·93
11. $M', \hat{M} \bar{y}$	40·4	54·8	0·88	42·8	59·7	0·87
12. $M'', \hat{M} \bar{y}$	36·6	54·8	0·77	38·4	59·7	0·81

$$R.B.\ (\hat{P}_{(2)}) = \frac{T-t}{Tt}[C_1^2 - \rho_3 C_1 C_3] \quad \ldots(13)$$

$$R.M.S.E.\ (\hat{P}_{(2)}) = \frac{T-t}{Tt}[C_1^2 + C_3(C_3 - 2\rho_3 C_1)]$$

$$+\frac{1}{Tt}\sum_{i=1}^{T}\frac{V(\hat{M}_i)}{\overline{M}^2} + \frac{1}{Tt}\sum_{i=1}^{T}\frac{M_i^2 V(\bar{y}_i)}{(\overline{M}\overline{Y})^2} \quad \ldots(14)$$

$$R.B.\ (\hat{P}_3) = \frac{T-t}{Tt}[C_1^2(1-\rho_1] + \frac{\frac{1}{T}\sum_{i=1}^{T}\overline{Y}_i - \overline{Y}}{\overline{Y}} \quad \ldots(15)$$

$$R.M.S.E.\ (\hat{P}_3) = \frac{T-t}{Tt}[2C_1^2(1-\rho_1) + C_2^2] + \frac{1}{T_t}\frac{\sum_{i=1}^{T} V(\hat{M}_i)}{\overline{M}^2}$$

$$+\frac{1}{Tt}\frac{\sum_{i=1}^{T} V(\bar{y}_i)}{Y^2} + \left[\frac{1}{T}\frac{\sum_{i=1}^{T}\overline{Y}_i - \overline{Y}}{\overline{Y}}\right]^2 \quad \ldots(16)$$

$$R.B.\ (\hat{P}_4) = \frac{T-t}{Tt}[C_1^2(3-2\rho_1) - \rho_3 C_1 C_3]$$

$$+\frac{1}{T}\frac{\sum_{i=1}^{T} M_i'\overline{Y}_i - \overline{M}'\overline{Y}}{\overline{M}'\overline{Y}} \quad \ldots(17)$$

$$R.M.S.E.(\hat{P}_4) = \frac{T-t}{Tt}[C_1^2(5-4\rho_1) + C_3^2 - 2\rho_3 C_1 C_3]$$

$$+\frac{1}{Tt}\frac{\sum_{i=1}^{T} V(\hat{M}_i)}{\overline{M}^2} + \frac{1}{Tt}\frac{\sum_{i=1}^{T} M_i'^2 V(\bar{y}_i)}{(\overline{M}'\overline{Y})^2}$$

$$+\left\{\frac{\frac{1}{T}\sum_{i=1}^{T} M_i'\overline{Y}_i - \overline{M}'\overline{Y}}{(\overline{M}'\overline{Y})^2}\right\}^2 \quad \ldots(18)$$

For estimating the relative bias and relative mean square error of these estimates of production, the estimates of ρ_1, ρ_2, ρ_3 and C_1, C_2 and C_3 have been obtained from the sample and substituted in the above formulae. In regard to the term of the form

$$\frac{\left(\frac{1}{T}\sum_{i=1}^{T} M_i''\overline{Y}_i - \overline{M}''\overline{Y}\right)^2}{(\overline{M}''\overline{Y})^2}$$

TABLE 3

Estimates of total milk production (000 tons) together with relative bias and relative mean square error

State	Category	Stratum	$\hat{P}_1=\hat{M}\times\frac{\sum_i^t \ddot{M}_i\bar{y}_i}{\sum_i^t \ddot{M}_i}$			$\hat{P}_2=\hat{M}\times\frac{\sum_i^t \hat{M}_i\bar{y}_i}{\sum_i^t \hat{M}_i}$			$\hat{P}_3=\hat{M}\times\frac{1}{t}\sum_i^t \bar{y}_i$			$\hat{P}_4=\hat{M}\times\frac{\sum_i^t \dot{M}_i\bar{y}_i}{\sum_i^t \dot{M}_i}$		
			Production	Rel. bias %	% S.E.	Production	Rel. bias %	% S.E.	Production	Rel. bias %	% S.E.	Production	Rel. bias %	% S.E.
Punjab	Cows	I	144	-0·7	7·35	145	@	6·81	146	0·2	6 82	145	-0 6	7·16
		II	192	-1·4	7·05	191	-0·3	6·14	186	-2·3	7·92	192	-0·4	6·15
		III	183	-0·6	3·79	186	-0·9	3·70	181	-4·9	7·06	185	-0·8	3 66
		Overall	519	-0·9	3·22	522	-0·4	3·22	513	-2·5	3·33	521	-0·6	3·29
	Buffaloes	I	478	0·1	1·70	477	-0·1	1·88	471	-1·6	2·39	480	0·4	2·05
		II	490	-5·2	13·76	492	-1·5	12·35	483	-11·8	17·79	490	-3·6	13·10
		III	632	0·3	6·25	649	0·1	6·19	593	-2·3	7·56	639	-0·4	6·05
		Overall	1600	-1·7	4·53	1618	-0·4	4·53	1547	-5·1	4·70	1609	-1·1	4·91
Eastern Uttar Pradesh	Cows	I	37	16·1	28·88	33	-2·4	26·32	31	-6·9	21·01	35	8·3	24·53
		II	60	2·3	16·95	58	-0·6	16·09	59	0·5	17·06	59	1·3	17·05
		III	77	2·7	14·08	75	@	13·39	79	4·5	17·37	78	4·5	15·62
		IV	48	-0·2	12·44	48	0·9	13·91	49	2·4	13·95	48	@	13·03
		Overall	222	4·2	8·20	214	-0·3	8·23	218	1·3	8·68	220	3·3	8·66
	Buffaloes	I	188	1·4	16·75	186	0.8	16·43	183	-1·1	15·94	191	2·8	16·25
		II	115	-0·7	12·95	115	1·0	13·66	116	1·3	13·04	115	0·2	12·95
		III	142	-0·4	19·89	145	1·5	18 00	156	9·3	19·84	146	2·7	19·35
		IV	137	-0·7	10·18	138	0·4	9·15	141	2·4	9·59	138	0·3	9·75
		Overall	582	0·3	7·72	584	0·9	7·72	596	2·9	7 88	590	1·7	7·97

@ Less than 0·1 per cent.

it can be seen that in view of the high correlation between $\hat{M}_i$'s and M_i'''s the contribution of this term to the relative bias and relative mean square error of $\hat{P}_{(1)}$ and $\hat{P}_{(4)}$ wil be small. In regard to $\hat{P}_{(3)}$, the estimate of the corresponding term will, however, be $(-\rho_2 C_1 C_2)$. Similarly, other terms, *viz.*.

$$\frac{1}{Tt} \sum_{i=1}^{T} \frac{V(\hat{M}_i)}{\bar{M}^2} \text{ and } \frac{1}{Tt} \sum_{i=1}^{T} \frac{M_i'' V(\bar{y}_i)}{(\bar{M}''\bar{Y})^2}$$

will be estimated from the sample.

The estimates of relative bias and relative mean square error of each of $\hat{P}_{(1)}$, $\hat{P}_{(2)}$, $\hat{P}_{(3)}$ and $\hat{P}_{(4)}$ are presented in Table-3.

5. Discussion

(i) *Estimate of number of animals in milk* :

In working out estimate of number of animals in milk in a stratum, three different estimates using ratio method of estimation were attempted. For the first two estimates, *viz.*, $\hat{M}_1$ and $\hat{M}_2$, the auxiliary variate used was the livestock census estimate of the number of animals in milk, while in regard to the third estimate, *viz.*, $\hat{M}_3$, the auxiliary variate was human population census estimate of number of households. The first estimate $(\hat{M}_1)$ differs from the second one in that a ratio was used at the first and second stages in case of the first one $(\hat{M}_1)$ whereas in case of second estimate $(\hat{M}_2)$, a simple estimate was taken at the second stage and a ratio was used at the first stage. The third estimate $(\hat{M}_3)$ was a ratio estimate at both the first and second stages. A study of the results presented in the Table 1 indicates that the second estimate $(\hat{M}_2)$ wherein a simple estimate was used at the second stage was much inferior to the other two estimates in the sense that it was subject to a higher mean square error and the relative bias was also large. Similarly, the third estimate obtained by using the number of households as auxiliary variate at both the stages had a higher mean square error than the one for which the auxiliary variate was the census estimate of number of animals in milk. This could be explained by the fact that the number of animals in milk in a tehsil (P.S.U.) during the season had a higher correlation with the livestock census estimate of the number of animals in milk than with the number of households in a tehsil. Thus the estimate ($\hat{M}_1$) for which a ratio method was used at both first and second stages and the auxiliary variate was the census estimate of number of animals in milk was preferred to the other two estimates and it had been used as one of the factors in the formula for estimating milk production.

(*ii*) *Estimate of total milk production* :

The estimate of total milk production in a stratum was obtained as a product of two factors, *viz* , the estimate of number of animals in milk and the average milk yield per animal in milk in the stratum. In order to study the relative efficiency of the different methods, four different estimates, *viz*., $\hat{P}_1$, $\hat{P}_2$, $\hat{P}_3$ and $\hat{P}_4$ were considered. For each of the estimates, the estimate of the number of animals in milk ($\hat{M}_1$) was the one obtained by using ratio at the first and second stages, and auxiliary variate was the livestock census estimate of the number of animals in milk. However, in regard to the second factor, *viz*., average milk yield per animal in milk per day, ($\overline{Y}$), four different estimates were considered. The first estimate of $\overline{Y}$ was the one obtained by weigting the tehsil averages with the census estimate of number of milch animals in the tehsil. The second estimate was obtained by weigting the tehsil averages by the survey estimate of the number of animals in milk in the tehsil. The third estimate of average was only a simple average of all the tehsil averages. The fourth estimate was obtained by using livestock census estimates of number of animals in milk as weights. In all the four cases, the tehsil average was the one defined earlier in Section 4.

A study of the results presented in Table 3 shows that the relative bias was of a higher order for the estimate $\hat{P}_3$ worked out by using simple mean of the tehsil averages. This estimate is also not consistent. Now the problem is to choose one among the other three estimates, *viz*., $\hat{P}_1$, $\hat{P}_2$ and $\hat{P}_4$. $\hat{P}_1$ and $\hat{P}_4$ as explained earlier, are of similar type in that the weights used for obtaining the stratum estimates of average milk yield were the census estimates of number of milch animals and animals in milk respectively. The estimate $\hat{P}_4$ is preferred to $\hat{P}_1$ since both the relative bias and relative mean square error of $\hat{P}_4$ were somewhat lower than those of $\hat{P}_1$. Moreover, $\hat{P}_4$ is more meaningful than $\hat{P}_1$. These two estimates as compared to $\hat{P}_2$ obtained by using the survey estimate of number of animals in milk as weight had a higher relative bias and relative mean square error. This was so since for $\hat{P}_2$ the weights were survey estimates worked out after taking into consideration the seasonal changes in the numbers. The mathematical expressions for the relative bias as also of relative mean square error of the $\hat{P}_2$ are also much simpler than those for any of the other three estimates. These results, therefore, indicate that $\hat{P}_2$ should be preferred to any of the other three estimates.

6. SUMMARY

The available official estimates of annual production of milk in the country are of limited utility since the data utilised were not those collected through any objective enquiries. The Institute of Agricultural Research Statistics had initiated pilot sampling enquiries during the Second Plan period in typical tracts of the different animal husbandry regions of the country with the object of evolving a suitable sampling technique for estimation of annual production of milk and study of bovine practices. This paper relates to the study of relative efficiency of different estimates considered for annual production of milk. The data collected from the surveys in Punjab and Eastern Uttar Pradesh had been utilised as a numerical example for the study. The estimate of total milk production in the tract was obtained as a product of the estimated number of animals in milk and the estimate of average milk yield per animal in milk in the tract. The results of the study indicated that for obtaining the estimate of number of animals in milk in a tract, ratio estimate using livestock census estimate of number of animals in milk as auxiliary variate at the first and second stages of sampling was the best. In regard to the second factor, *viz.*, average milk yield per day per animal in milk, four different estimates were considered. In the four estimates, the estimate of the average milk yield at tehsil level was the same, while for working out the stratum estimate, different types of weights were used. The different weights used were the census estimate of milch animals, animals in milk, survey estimate of number of animals in milk and equal weights. A study of the results indicated that for estimating total milk production in the stratum, the estimate of stratum average milk yield per animal in milk should be the one obtained by weighting tehsil averages with the estimated number of animals in milk in the tehsil.

7. Acknowledgements

The authors are thankful to Sarvashri O. P. Jain and K. B. Singh who helped in carrying out the necessary computations for the numerical example.

REFERENCES

1. Amble V.N. et al. (1965) : Milk production of Bovines in India and their feed availability, Ind. J. of Vet. Sci. Vol. 35, pp. 221-238.

2. Chowdhary, R K. and Narang, K.L. (1962) : Milk production, its technique of estimation—Agricultural Situation in India, Vol. XVII, No. 2, pp. 124-128.

3. Panse, V.G. & et al. (1964) : Sample Survey for estimation of Milk production (Punjab)—ICAR Report Series.

4. ——————(1966) : Sample Survey for estimation of Milk production (Eastern U.P.)—I.C.A.R. Report Series No. 24.

5. Singh, D. and Murty, V.V.R. (1963) : Random sample survey technique for estimation of production and consumption of milk—Agricultural Situation in India, Vol. XVIII, No. 1.

6. Sukhatme, P.V. (1953) : Sampling Theory of surveys with applications.