EFFICIENCY OF STRATIFICATION IN LIVESTOCK SURVEYS*

By

R.C. Garg** and V.V.R. Murty

Institute of Agricultural Research Statistics, New Delhi

(*Received in April, 1971*)

Introduction

The Institute of Agricultural Research Statistics had conducted a series of pilot investigations for evolving suitable sampling techniques for improvement of statistics of livestock numbers and their products. For the investigations conducted on livestock products, stratification usually adopted was on the basis of geographical contiguity. In this paper, this method of stratification using geographical contiguity has been compared with alternative methods of stratification. In brief, the problems considered here are :

(*a*) To study how far the geographical stratification has reduced the variances of estimates of number of milch cows and annual milk production, the characters under study.

(*b*) Alternative methods of construction of strata on the basis of an auxiliary character. This study includes determining number of strata and strata boundaries.

(*c*) Efficiency of different methods of stratification and different methods of estimation as compared to simple random sampling (SRS).

---

*Part of M.Sc. thesis submitted by the first author to the Post Graduate School, IARI (1967).

**Present address : Junior Programmer, Indian Veterinary Research Institute, Izatnagar, U.P.

The various methods of construction of strata examined are:

(*a*) Equal aggregate output method.
(*b*) Equal interval on cumulative $\sqrt{f}$ method.
(*c*) Ekman's method.

These problems have been examined with the help of a numerical example using the data on number of milch cows and milk production collected from the sample survey conducted in Kerala State during 1964-65. The auxiliary variate which was highly correlated with the characters under study has been used as a stratification variable. The auxiliary variate chosen for stratification is number of milch cows (X) as per livestock census, 1961. The correlation coefficient between X and enumerated number of milch cows (Y) from the survey is 0.96 and between X and milk production (Z) is 0.76.

2. Design of the survey adopted in Kerala State

The whole State was divided into three natural regions, *viz.* lowland, mid-land, and high-land. The cross-section of a district with these regions was taken as a stratum. In all, 14 strata were thus formed. A cluster of two villages (one selected at random and the other nearest to it) was the primary sampling unit. From each stratum, six to nine clusters of villages were selected during the year. A cluster of three adjacent households was the unit at the second stage of sampling. At the third stage, two animals in milk were selected from each household for recording milk yield. In each cluster of villages, a complete enumeration of animals in milk and also milch animals was carried out during the first month of a season. Data on a day's milk yield of the sample of animals in milk selected from each household as also information on the feeding and management practices of all cattle and buffaloes in the household were collected. The total number of clusters of villages selected during the year was 90.

3. Efficiency of Geographical Stratification

In order to study whether geographical stratification used in the survey has reduced the variation in estimates of the number of milch cows and milk production, the mean squares within (s^2_w) and between (s^2_b) strata have been calculated from the data by the technique of analysis of variance. The complete analysis of variance table is presented below :

Source of Variation	*D.F.*	*S.S.*	*M.S.*	*Expected value of M.S.*
Between strata	$L-1$	$\sum_{h=1}^{L} n_h(\bar{y}n_h-\bar{y}_n)^2$	s^2_b	$\sum_{h=1}^{L}\sigma^2_h(n-n_h)/(nL-n)$ $+\sum_{h=1}^{L} n_h\left(\mu_h-\frac{\sum_{h=1}^{L} n_h\mu_h}{n}\right)^2/(L-1)$
Within strata	$n-L$	$\sum_{h=1}^{L}\sum_{i=1}^{n_h}(y_{hi}-\bar{y}_{nh})^2$	s^2_w	$\sum_{h=1}^{L}\sigma^2_h(n_h-1)/(n-L)$
Total	$n-1$			

where

L=Number of strata,

$n=\sum_{h=1}^{L} n_h$=total sample size, n_h being the sample size in h-th stratum,

y_{hi}=value of i-th unit in h-th stratum,

$\bar{y}_{nh}$=sample mean of h-th stratum,

$\bar{y}_n$=grand sample mean,

and the model used is :

$$y_{hi}=\mu+\mu_h+e_{hi},\ i=1, 2...n_h;\ h=1, 2\ ..L,$$

where

μ=grand mean of the population

μ_h=effect of h-th stratum

and e_{hi} is a random variable with $E(^ehi/h)=0$; $V(e_{hi}/h)=\sigma^2h$

Looking at the expected value of the two mean squares, it is seen that unless σ^2h is constant or n_h is constant, no conclusion about the efficiency of stratification by the ratio of between strata mean square to within strata mean square can be drawn. However, test of the ratio can roughly indicate whether μ_h, ($h=1, 2.....L$) are

differing significantly or not. If μ_h's are differing, the present stratification may lead to gain in efficiency.

Tables 1 and 2 below give analysis of variance for the two characters.

TABLE 1

Analysis of variance of number of milch cows

Source of variation	*D.F.*	*S.S.*	*M.S.*	*F*
Between strata	13	321978	24767	1.927
Within strata	76	976509	12849	
Total	89	1298487		

It is seen from the results of Table 1 that '*F*' value is not significant at 1% level though it is just significant at 5% level. This shows that μ_h's may not be differing much for the geographical stratification used under the survey. However, this may have to be confirmed from further studies.

TABLE 2

Analysis of variance of milk production $(Kg)^2$

Source of variation	*D.F.*	*S.S.*	*M.S.*	*F*
Between strata	13	143954	11073	2.505
Within strata	76	335960	4420	
Total	89	479914		

It is seen from Table 2 that '*F*' value is significant at 1% level. This indicates that μ_h's are considerably different from one another and may lead to gain in efficiency.

Alternative methods of stratification have been considered for both the characters. In view of the fact that the correlation between census estimate of number of milch cows and each of the two characters considered is high (viz. 0.96 and 0.76), the census estimate has been utilised as the auxiliary variate for the construction of strata.

4. Construction of Strata According to Criteria other than Geographic Contiguity

Three methods of construction of strata have been considered and compared. These are :

(*a*) Equal aggregate output method due to Mahalanobis (1952).

(*b*) Equal interval on cumulative $\sqrt{f}$ method proposed by Dalenius and Hodges (1957) where f is the frequency function.

(*c*) $W_h\,(X_h - X_{h-1}) = \text{Constant}$ by Ekman (1959) where W_h is the total frequency in stratum h and X_h, X_{h-1} are the upper and lower boundary point of stratum h.

The efficiency of the above three rules has been studied on the basis of the survey conducted in Kerala State during 1964-65.

(*a*) *Equal aggregate output method* :

Using this method, the strata boundaries have been worked out from the data and are shown in Table 3.

TABLE 3

Strata boundaries for varying number of strata (Using census estimate of milch cows as per Livestock census, 1961 as the stratification variable)

Number of Strata	*Stratum* 1	2	3	4
$L=2$	0–217	>217		
$L=3$	0–174	174–304	>304	
$L=4$	0–130	130–217	217–347	>347

(*b*) *Equal interval on cumulative* $\sqrt{f}$ *method* :

The strata boundaries obtained by this procedure are indicated in Table 4.

TABLE 4

Strata boundaries for varying number of strata (Using census estimate of milch cows as per census 1961, as the stratification variable)

No. of strata	*Stratum* 1	2	3	4
$L=2$	0–174	>174		
$L=3$	0–87	87–217	>217	
$L=4$	0–87	87–174	174–260	>260

(c) *Ekman's Method* :

In this method, the product of the cumulative frequency within the stratum and the width of the stratum is equalized. The first step is to cumulate 'f' values. This method is somewhat complicated and difficult to apply. A rough guide suggested by Ekman is to compute the product $Q=W\times R$ where R is the range of X. Then the constant value per stratum is approximately given by Q/L^2 where L denotes the number of strata. This relationship is exact for a rectangular distribution but tends to give the high results for a highly skewed distribution. The strata boundaries obtained by this procedure are indicated in Table 5.

TABLE 5

Strata boundaries for varying number of strata (using census estimate of number of milch cows as per census, 1961, as the stratification variable)

No. of strata	*Stratum* 1	2	3	4
$L=2$	0–217	>217		
$L=3$	0–130	130–304	>304	
$L=4$	0– 87	87–174	174–347	>347

5. Discussion and Conclusions

Tables 6 and 7 below give the relative efficiency of the estimated number of milch cows and annual production of cow milk for different methods of construction of strata using proportional allocation. For the purpose of a comparative study, relative efficiencies

TABLE 6

Relative efficiency of different methods of construction of strata for the character-number of milch cows

Method of estimation	*Method of stratification*	*Equal aggregate output method*			*Equal interval on cum. $\sqrt{f}$ method*			*Ekman's method*			*Design used under survey*	*SRS*
		No. of strata			*No. of strata*			*No. of strata*				
		2	3	4	2	3	4	2	3	4		
Simple estimate		195	237	271	191	232	258	195	246	269	105	100
Ratio estimate		1324	1268	1251	1333	1324	1295	1324	1277	1259	981	1315
Regression estimate		1395	1634	1727	1374	1384	1474	1395	1634	1744	1296	1374

TABLE 7

Relative efficiency of different methods of construction of strata for the character-milk production of cows

Method of estimation	*Method of stratification*	*Equal aggregate output method*			*Equal interval on cum. $\sqrt{f}$ method*			*Ekman's method*			*Design used under survey*	*SRS*
		No. of strata			*No. of strata*			*No. of strata*				
		2	3	4	2	3	4	2	3	4		
Simple estimate		130	133	139	133	135	135	130	132	141	158	100
Ratio estimate		222	220	228	228	227	225	222	229	224	417	233
Regression estimate		262	280	281	262	266	278	262	282	280	859	247

for SRS and the geographical stratification have also been presented. In working out the relative efficiencies, the variance of the simple estimate obtained from SRS has been taken as 100.

(a) Relative efficiency for estimates of number of milch cows

It is seen from Table 6 that stratification using census estimate of number of milch cows (X) which is highly correlated with the enumerated number (Y) has helped in substantial reduction in the sampling variance of the estimate of number of milch cows. Again the ratio and regression methods of estimation, used with such stratification, have further lowered down the sampling error of the estimate.

It is seen from Table 6 that the relative efficiency of three methods of stratification for different methods of estimation is large as compared to that of the design used in the survey for the same methods of estimation. It is thus concluded that geographical stratification is not as efficient as alternative methods of stratification for the estimation of number of milch cows.

It is also seen from Table 6 that the equal aggregate output method and Ekman's method of stratification gave almost equally efficient estimates. The efficiency obtained by stratifying units according to cumulative $\sqrt{f}$ method and by using simple estimate or regression estimate is higher than the corresponding estimates obtained by the other two methods of stratification. However, the efficiency of the ratio estimate with stratification based on cumulative $\sqrt{f}$ is the lowest. It may thus be concluded that the equal aggregate output method and Ekman's method may prove more efficient than the cumulative $\sqrt{f}$ method.

It may, however, be seen from Table 6 that if SRS accompanied by ratio-estimate was used, the efficiency so obtained is not further improved by adopting stratification. But stratification is to be adopted for administrative convenience as well as for estimating annual production of milk where there is considerable scope for proper stratification to reduce the sampling errors as will be seen while discussing the results presented in Table 7.

(*b*) *Relative efficiency for the estimates of annual milk production*

The results presented in Table 7 indicate that the relative efficiencies for the three methods of stratification for different methods of estimation are almost of the same order. Further the relative efficiency for the design used for the simple estimate is 158 and it increases to 417 and 859 when ratio and regression methods of estimation were used. These relative efficiencies of the design used are quite large as compared to those of the three methods of stratification. The significance of these results is that geographical stratification is more efficient than the alternative methods of construction of strata using auxiliary variate and SRS for the character-milk production of cows.

SUMMARY

Stratification adopted in livestock Surveys is examined regarding its efficiency in reducing sampling errors of estimates of the number of milch cows and the milk production. Alternative methods of construction of strata on the basis of an auxiliary character are also studied to see their efficiency in comparison to the geographical stratification. In addition, the use of ratio and regression techniques coupled with stratification of various types is investigated to assess further reduction in variance of estimates. The data collected on milch cows in the sample survey for the estimation of milk production in Kerala State formed the basis of this paper.

ACKNOWLEDGEMENTS

We wish to record our deep sense of gratitude to Dr. G.R. Seth, Ex. Director, I.A.R.S., New Delhi (now on FAO assignment) under whose guidance this work was done. Our thanks are also due to the referee for his valuable comments which contributed in further improving this paper.

REFERENCES

Cochran, W.G. (1953). : Sampling Techniques, John Wiley and Sons Ltd., New York.

Cochran, W.G. (1961) : *Comparison of Methods for Determining Stratum Boundaries*, Bulletin of the International Statistitical Institute, Vol. 38, No. 2.

Dalenius, T. (1950). : *The Problem of Optimum Stratification.* Skand. Akt : 203-13.

Dalenius, T. & Hodges, J.L. (1959). : *Minimum Variance Stratification*, Journal of the American Statistical Association, Vol. 54, 88-101.

Ekman, G. (1959). : *An Approximation Useful in Univariate Stratification.* The annals of Mathematical Statistics, Vol. 30, 219-29.

Mahalanobis, P.C. (1952). : *Some Aspects of the Design of Sample Survey*, Sankhya, Vol. 12, 1-7.