Bulletin of the Rubber Research Institute of Sri Lanka, (1998) 37, 75-82

OPTIMUM SAMPLE SIZE FOR RUBBER SECTOR SOCIO ECONOMIC STUDIES

P H M U Herath

ABSTRACT

This study indicates that the rubber sector could be divided intensively into non overlapping groups. The division is tally with the sector performances as well as the administrative structure. Therefore, stratified random sampling procedure would be employed effectively to obtain representative data of the rubber growers population for socio economic studies. The study describes the sampling procedure for the different strata.

INTRODUCTION

The modern statistics is a theory of information with inference as its objective. The medium of inference is the sample, which is a subset of measurements selected from the population. We wish to make an inference about the population based on the characteristics of the sample.

Each observation or item, taken from the population, contains a certain amount of information about the population parameters of interest. Since information costs money, we must determine how much information we should buy. Too little information prevents us from making good estimates, while too much information results in a waste of money.

The quantity of information obtained in a sample depends upon the number of items sampled and upon the amount of variation in the data. Sample variation can be controlled by the method of selecting the sample called the design of the sample survey; it, along with the sample size, determines the quantity of information in a sample pertinent to a population parameter.

This study focused on economics of purchasing a specific quantity of information. More specifically, design a sample procedure that reduce the cost of a fixed quantity of information.

Present structure of the rubber sector

In Sri Lanka rubber is planted in commercial level in 14 districts as indicted in the figure 1. The area under rubber and the number of farmers (fig. 2) vary in different districts. Also, the rubber growers could be divided into two main sectors

namely, estate and smallholder sector. These two sectors could be further demarcated based on whether they bear immature, mature or both immature and mature rubber together. Further, these holdings could be separated into different strata according to rubber clone types and finally whether they intercropped or not.

area (ha)

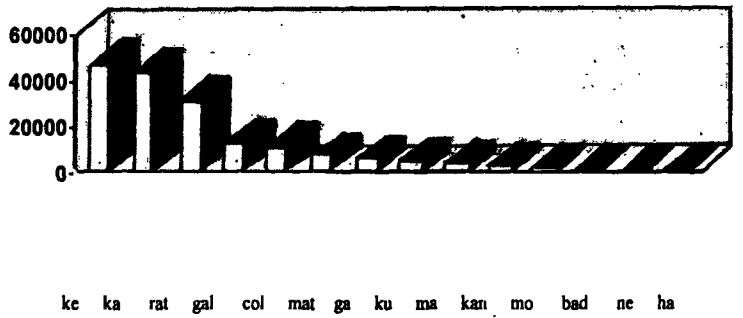


Fig 1. Rubber area distribution by districts

Source: Census agricultural crops and livestocks 1992/93

number (000)

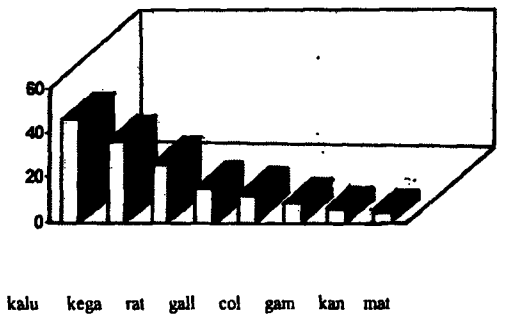
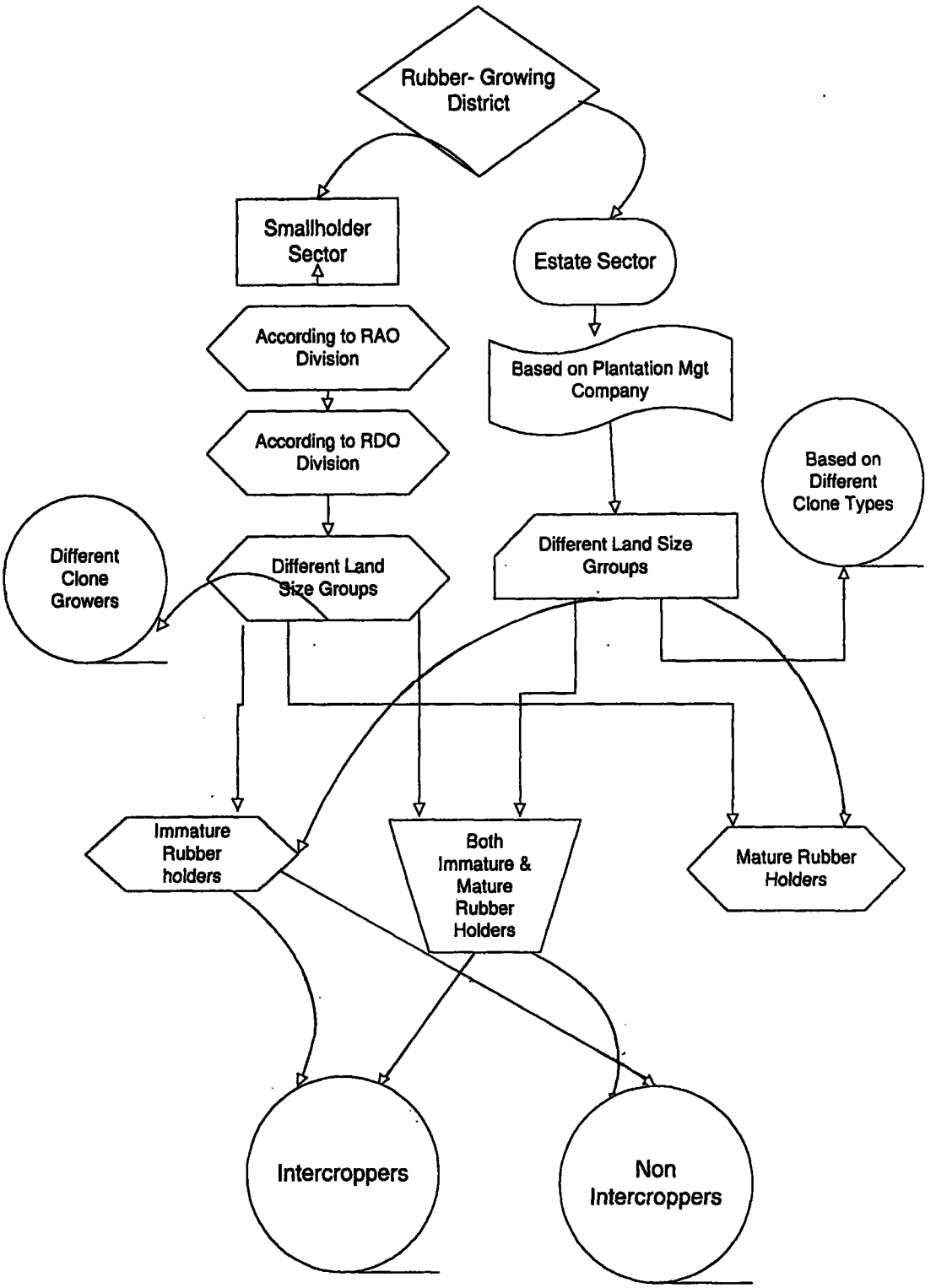


Fig 2. Number of smallholders by districts

source : Rubber Control Department

According to the above information the rubber sector could be clearly demarcated into different strata as indicated in the figure.



Differentiation of strata for the study population

According to the figure (1), for socio economic studies, the rubber sector population could be divided into 14 main strata, *i.e.* district basis demarcation. Main strata could differentiate into sub strata depending on the nature of the study. The demarcation of the strata is justified by acceptable criteria like administrative or rubber sector specific factors. In this article an example was discussed on sampling of smallholder rubber growers population, to study the crop diversification of rubber sector.

Sampling procedure

Based on the above information the stratified random sampling procedure was identified as one of the best method of sampling for socio economic studies on rubber. Kegalle, Kalutara, Rathnapura, Colombo and Gampaha districts were selected out of 14 main strata. These 5 districts are covering around 80 percent of the rubber growers population. Therefore, we could consider these districts as representative districts of the rubber growers. According to the Rubber Development Department administration procedure Ratnapura, Kalutara, Kegalle, Colombo and Gampaha districts were divided into 31, 45, 54, 24 and 15 Rubber Development Officers (RDO) ranges respectively. Rubber activities of each range is conducted by RDO. Therefore, extension activities, subsidy distribution etc. are unique for each range. So, these ranges could be considered as sub strata with small within variability in each stratum. However, for sampling it is very costly to cover each and every RDO range, as intensive traveling is necessary. Therefore, to select the optimum number of RDO ranges out of the total 156 ranges the following formula was used.

$$n = \frac{{}_{i=1}\Sigma^{L} N_i^2 \sigma_i^2 / W_i}{N^2 D + {}_{i=1}\Sigma^{L} N_i \sigma_i^2} \qquad (1)$$

$$D = B^2/4N^2$$

where,

N - total number of RDO ranges

N_i - number of RDO ranges in each district

σ_i - yield variance in RDO ranges of particular district (approximated by sample variance of previous experiments)

W_i = fraction of total sample size to be allocated into i^{th} stratum

B = specified amount by experimenter to lie the estimates within that unit of population mean with probability equal to 95.

There was no previous information of intercropping studies on rubber. Therefore, to estimate the parameters specified for the above equation, the information of the previous studies on some other aspects was used. This procedure was justified because the final objective was to differentiate the concerned population into homogenous groups. Therefore, a study on yield performance was used as a base to estimate the parameters specified in the above equation.

Accordingly,
$\sigma = 50$
B = 1000 (in the case of estimating total yield figure)

The W_i the fraction of the total ranges to be allocated to each district was estimated by the following formula.

$$Wi = \frac{N_i\sigma_i / \sqrt{C_i}}{\sum_{i=1}^{L} N_i\sigma_i / \sqrt{C_i}} \qquad (2)$$

where,
N_i = number of RDO ranges in relevant districts.
C_i = cost of obtaining information from a farmer (in Kalutara Rs. 125 per farmer and 150 per farmer in other districts)

The W_i values were estimated by the formula 2 as follows and substituted to the formula 1 to estimate the optimum number of RDO ranges to be selected.

W_i (rathn) = (31*50)/√150/ (31*50)/√150+ (45*50)/√125+ (56*50)/√150+ (24*50)√150 = 0.19

W_i(kalutara) = (45*50)/√125/ (31*50)/√150+ (45*50)/√125+ (56*50)√150+ (24*50)/√150 = 0.28

W_i(kegalle)= (56*50)/√150/ (31*50)/√150+ (45*50)/√125+ (56*50)/√150+ (24*50)/√150 = 0.35

W_i(colo) = (24*50)/√150/ (31*50)/√150+ (45*50)/√125+ (56*50)/√150+ (24*50)/√150 = 0.15

by substituting the above Wi values to the formula 01

$$n = \frac{{}_{i=1}\Sigma^{L} N^2\sigma_i^2/W_i}{N^2D + {}_{i=1}\Sigma^{L} N_i\sigma_i^2} \qquad (1)$$

${}_{i=1}\Sigma^{L} N^2\sigma_i^2/W_i$
$= (31)^2.(50)^2 / 0.19 + (45)^2.(50)^2/0.28 + (56)^2.(50)^2 / 0.35 + (24)^2.(50)^2/ 0.18$
$= 61125094$

$N^2D = 250000$ as $D = B^2/4N^2$

${}_{i=1}\Sigma^{L} N_i\sigma_i^2$
$= 31*(50)^2 + 45*(50)^2 + 56*(50)^2 + 24*(50)^2 = 390000$
$n = 61125094/(250000 + 390000) = 96$

According to the above calculation the total number of RDO ranges was estimated as 96.

By using the above information the optimum number of RDO ranges for each district was estimated by the following formula,

$$n_i = n \frac{N_i\sigma_i/\sqrt{C_i}}{{}_{i=1}\Sigma^{L} N_i\sigma_i/C_i}$$

where, n = 96

$n_{(rathn)} = 96 *(31*50)/\sqrt{150}/(31*50)/\sqrt{150} + (45*50)/\sqrt{125} + (56*50)/\sqrt{150} + (24*50)/\sqrt{150} = 19$

$n_{(kalut)} = 96 *(45*50)/\sqrt{125}/(31*50)/\sqrt{150} + (45*50)/\sqrt{125} + (56*50)/\sqrt{150} + (24*50)/\sqrt{150} = 28$

$n_{(kegalle)} = 96* (56*50)/\sqrt{150}/ (31*50)/\sqrt{150} + (45*50)/\sqrt{125} + (56*50)/\sqrt{150} + (24*50)/\sqrt{150} = 34$

$n_{(colo)} = 96* (24*50)/\sqrt{150}/ (31*50)/\sqrt{150} + (45*50)/\sqrt{125} + (56*50)/\sqrt{150} + (24*50)/\sqrt{150} = 15$

Sample sizes for selected RDO ranges

The total number of farmers to be interviewed from each district was decided by considering the cost factor. Calculations have showed an average cost of Rs. 125 in Kalutara and Rs.150 per farmer in the rest of the districts. Assumed that the total estimated budget for the survey is Rs.100000 and this amount is to be spent equally among the districts. Therefore, number of farmers to be interviewed is 200 from Kalutara district and 168 from each of the other districts.

Assuming equal cost for each RDO range and the variances are equal in each RDO range of a given district the formula 2 was simplify as follows to estimate allocation fractions for each RDO range.

$n_{RDO} = n * N_i / N$

where,
N = total number of immature rubber holders in a particular district. (immature rubber growers population was selected because only they can do intercropping).

N_i = number of immature rubber holders in a particular RDO range.
n = number of farmers to be interviewed in a district.

According to the RDD sources in Rathnapura district N= 2343 and number of immature rubber holders vary in different RDO ranges. Therefore, number of farmers to be allocated for different RDO ranges could be calculated as follows.

$n_{rat} = 168* (101/2343) = 07$
$n_{ellawea} = 168 * (264/2343) = 19$
$n_{karadena} = 168 * 142/2343 = 10$ etc.

The above method of estimating sample sizes to a strata is called proportional allocation.

The same procedure was employed for further categorization of the number of farmers to be included to different holding sizes. Accordingly,

$n_{land\ size} = n* N_i/N$
where,
N_i = number of immature farmers in given land category of a particular RDO range
N = number of immature farmers in particular RDO range.
n = number of farmers to be interviewed from a given RDO range

for example,

in RDO range ellawela of the Rathnapura district the number of immature rubber holders according to the different land sizes is as follows

less than 5 acres = 254, 5-10 acres = 9, 10-25 acres = 3, 25-50 acres=0,

Therefore,

$n_{<5ac} = 19 * (254/264) = 16$

$n_{5\text{-}10acres} = 19 * (9/264) = 02$ etc.

CONCLUSION

1. Selecting optimum sample size for socioeconomic studies is a critical factor as the sample is supposed to represent the concerned population characters.
2. According to well documented information the rubber sector could be clearly divided into non overlapping groups called strata.
3. The potential of stratification of the sector is very intensive. Therefore, homogenous information could be gathered from each stratum.
4. High homogeneity in sampling group increased the quantity of information gathered. Also, the cost of conducting actual sampling tends to be less because of administrative convenience.
5. According to the above information stratified random sampling is effective for socio economic studies of rubber sector, because the method is comply with the structure of the sector as well as certify maximum information for minimum cost.

REFERENCES

Administrative Reports, Rubber Development Department, Various issues.

Cochran, W G (1963). *Sampling Techniques*, 2d ed., New York; John Willey & Sons, Inc., 1963.

Plantation Sector Statistical Pocket Book (1995). Planning Division. Ministry of Public Administration Parliamentary Affairs and Plantation Industries.