OPTIMUM STRATIFICATION FOR YIELD ESTIMATION IN COCOA (*Theobroma cacao* L.)

By

SUNILKUMAR. G.

THESIS

Submitted in partial fulfilment of the requirement for the degree of

Master of Science in Agricultural Statistics

Faculty of Agriculture

Kerala Agricultural University

Department of Agricultural Statistics

COLLEGE OF HORTICULTURE

VELLANIKKARA, THRISSUR - 680 656

KERALA, INDIA

2000

DECLARATION

I hereby declare that this thesis entitled **"OPTIMUM STRATIFICATION FOR YIELD ESTIMATION IN COCOA (*Theobroma cacao.*L)"** is a bonafide record of research work done by me during the course of research and that this thesis has not previously formed the basis for the award to me of any degree, diploma, associateship, fellowship or other similar title of any other university or society.

College of Horticulture
Vellanikkara
Date: 27/04/2000

Sunilkumar,G
(97-19-02)

Dr. V. K. Gopinathan Unnithan
Associate Professor

Department of Agrl. Statistics
College of Horticulture
Kerala Agricultural University
Vellanikkara, Thrissur.

CERTIFICATE

Certified that this thesis entitled **"OPTIMUM STRATIFICATION FOR YIELD ESTIMATION IN COCOA (*Theobroma cacao.*L)"** is a record of research work done by Mr. Sunilkumar,G under my supervision and guidance and that it has not previously formed the basis for the award of any degree, fellowship or associateship to him.

Dr. V. K. Gopinathan Unnithan
(Chairman, Advisory Committee)

Place: Vellanikkara
Date: 27/06/2000

CERTIFICATE

We, the undersigned members of the Advisory Committee of Mr. Sunilkumar,G , a candidate for the degree of the Master of Science in Agricultural Statistics, agree that this thesis entitled **"OPTIMUM STRATIFICATION FOR YIELD ESTIMATION IN COCOA (*Theobroma cacao.*L)"** may be submitted by Mr. Sunilkumar,G in partial fulfilment of the requirement for the degree.

Dr. V. K. Gopinathan Unnithan
(Chairman, Advisory Committee)
Associate Professor
Department of Agrl. Statistics
College of Horticulture
Kerala Agricultural University
Vellanikkara, Thrissur

Dr. P. V. Prabhakaran
(Member, Advisory Committee)
Associate Dean (i/c)
Professor and Head
Department of Agrl. Statistics
College of Horticulture
Kerala Agricultural University
Vellanikkara, Thrissur

Dr. R. Vikraman Nair
(Member, Advisory Committee)
Associate Director of Research
Southern zone
College of Agriculture
Vellayani, Thiruvananthapuram

Smt. Gracemma Kurien
(Member, Advisory Committee)
Assistant Professor
Department of Agrl. Statistics
College of Horticulture
Kerala Agricultural University
Vellanikkara, Thrissur

Approved
(Dr. V. K. R. Nair)
EXTERNAL EXAMINER

Acknowledgment

I am deeply indebted to **Dr. V. K. Gopinathan Unnithan**, Associate Professor, Department of Agricultural Statistics, College of Horticulture and Chairman of my advisory committee for his constructive criticism, valuable advice and encouragement throughout the course of this investigation. He not only supervised my work but also was the motivating force behind my efforts for completing the course. I consider myself fortunate in having had the privilege of being guided by him.

Dr. P. V. Prabhakaran, Associate Dean (i/c) and Professor and Head, Department of Agricultural Statistics, College of Horticulture and member of my Advisory committee guided me with his technical advice. I sincerely thank him for having devoted his valuable time in dextrous guidance and learned counsel.

Smt. Gracemma Kurien, Assistant Professor, Department of Agricultural Statistics, College of Horticulture and member of my Advisory committee provided solid and timely support at all stages of the thesis preparation.

I take this opportunity to express my indebtedness to **Dr. R. Vikraman Nair**, Associate Director of Research, Southern zone, College of Agriculture, Vellayani and member of my advisory committee for providing the necessary data and for the valuable suggestions rendered by him.

I wish to acknowledge the assistance and advice provided by **Shri. S. Krishnan, Smt. P. Soudamnini, Smt. T.K. Indira Bai and Smt. T.K.Ajitha,** Assistant Professor, Department of Agricultural Statistics and **Dr. V.K.Mallika**, Associate Professor, CCRP, College of Horticulture during the course of my project work.

I would like to keep on my record my deep sense of gratitude to Smt. Joyce, Technical Assistant, College of Horticulture and research assistants of CCRP for their sincere help and co-operation.

I wish to express my gratitude to my seniors Manoj Kumar and Sree Kumar, my classmate Priyalashmi and my juniors Sajitha, Shiji, Sangeetha and Mishra without whom the completion of this thesis would have been difficult.

I had the privilege to have the sincere help from all corners, from my student friends, both post and under graduates. More personally I would like to thank the sincere help and inspiration from Murali, Ajith (K, C.B, and Kumar), Mash, Sainu, Sunil, Sherin, Pattabi, Subash, Prakash, Dutt, Reddy, Sajna, Renjith, Biju, Sanjeev, Shibu, Rajeev, Tony, Krishnakumar, Sreeja, Annie and Anuchechi.

The award of the Kerala Agricultural University Junior fellowship is gratefully acknowledged.

No words can explain the affection and gratitude to my beloved parents and brothers who provided me the soothing shoulders in time of need.

Above all, I bow before God, the almighty for his bountiful blessings without which I could not have completed this endeavour successfully.

Sunilkumar, G

To

My Beloved Parents

CONTENTS

LIST OF TABLES

List of tables (continued)

LIST OF APPENDICES

INTRODUCTION

INTRODUCTION

Different sampling techniques aim at estimating means of one or more population characteristics. The precision of an estimator of the population mean depends not only on the size of the sample and the sampling fraction but also on the variability or heterogeneity present in the population. Simple random sampling makes use of information on the population only to the extend of identity of units. Hence, sampling procedures that make use of the additional information about the population are developed to increase precision of the estimator. One such procedure is stratified random sampling, which consists in dividing the population into different classes known as 'strata' based on the available information about the population and drawing random samples from each stratum, which together will form the required sample.

Stratified random sampling has tremendous advantages over simple random sampling. It makes essential background for good representation of population in the sample and thereby efficiency of estimation increases, particularly when there is considerable variability in the population.

Efficiency of the estimator of the population mean from a stratified simple random sample depends on

- the number of strata
- the relevant stratification variable / variables
- the boundary points of strata or the method of constructing strata : and
- the type of allocation.

For allocating samples to different strata, techniques have been developed successfully. Equal, Proportional, Neyman and Optimum allocations of sample size are well known solutions. The problem of choosing stratification variable/variables needs attention. It depends on the availability of information on auxiliary character or characters that have a close relation with the study character or characters. Determination of the number of strata into which the population is to be divided is an area where serious attempts have yet to be made. The remaining problem is the choice of boundary points of strata. Various workers proposed various methods to determine approximately optimum boundary points of strata as the formulae for optimal stratification is not easily tractable. The present study aims at comparing some of those methods that are adjudged good by one or more authors and at selecting a quick method of determining strata boundaries from among the available methods of optimum stratification for estimating the yield of cocoa (*Theobroma cacao*.L).

Cocoa (*Theobroma cacao*.L) is a perennial crop that gains great importance, especially among farmers of Kerala. It belongs to the family 'Sterculiaceae' and was originated in Brazil. Both Criollo and Forestro types are cultivated in India, mainly as an intercrop in coconut and arecanut gardens, due to its shade loving nature. The crop is grown mostly in Kerala and Karnataka states in India and Kerala contributes about 75 percent of the total production of the country (Nampoothiri and Balasimha, 1999). The total area under the crop is around 12,000 hectares in Kerala and 14,618 hectares in India. The corresponding production in metric tonnes are 5,000 and 7,837 respectively (Balasubramanian, 1999). The projected area and production for 2002 are 24,000 hectares and 20,000 tonnes respectively for India. The requirement for domestic industry would be about 16,000 to 20,000 tonnes as against the present grinding capacity of 9,250 tonnes. Thus the demand for cocoa is increasing and hence is very important in Kerala economy in general and to the farmers in particular. Hence research work on the crop is to be taken up extensively.

If an appropriate scheme of stratification of cocoa trees for estimation of future yield is arrived at, yield prediction can be done efficiently. More over the technique can be made use of advantageously for blocking trees for conducting experiments in the crop.

Therefore the present investigation was taken up with the following objectives.

1. To compare different rules of stratification on the basis of different auxiliary characters for estimating the yield of cocoa.

2. To select a quick method for determining the strata boundaries for yield estimation in cocoa.

3. To select appropriate variable / variables that can be used for stratification for yield estimation in cocoa.

REVIEW OF LITERATURE

REVIEW OF LITERATURE

Very few works have so far been done on determination of optimum strata boundaries either in univariate or multivariate case. A brief account of work done in methodology as well as application is given in three sections of this chapter, viz, stratification on the study character itself, stratification on an auxiliary character and multivariate approach.

2.1 Study character as the stratification variable

Almost all major works on the determination of optimum strata boundaries in the beginning was done using the study variable itself as the stratification variable. A review of some major works done by taking the study variable as the stratification variable is presented in this section.

Dalenius (1950) derived the necessity conditions to be satisfied by the boundary points of strata in order to achieve minimum variance of the estimator of population mean based on a stratified simple random sample under proportional and Neyman allocations of sample size, assuming stratification on study character itself. The general solution for proportional allocation was

$$X_h = \frac{(\mu_h + \mu_{h+1})}{2} \qquad h=1,2,3,\text{---------},L-1$$

while, for Neyman allocation, it was

$$\frac{\sigma_h^{\,2} + (x_h - \mu_h)^2}{\sigma_h} = \frac{\sigma_{h+1}^{\,2} + (x_h - \mu_{h+1})^2}{\sigma_{h+1}}$$

$$h = 1,2,3,\text{---------},L-1.$$

where μ_h the mean of the h^{th} stratum

σ_h^2 the variance of the h^{th} stratum

x_h the boundary point between h^{th} and $h+1^{st}$ strata: and

W_h the relative size of the h^{th} stratum.

Raj (1968) derived the equations to be satisfied by the boundary points of strata under equal allocation to be

$$W_h\left(\sigma_h^2 + (x_h - \mu_h)^2\right) = W_{h+1}\left(\sigma_{h+1}^2 + (x_h - \mu_{h+1})^2\right),$$

$$h = 1,2,3,\text{---------},L\text{-}1$$

Since these solutions are not easy to achieve, several workers proposed approximate solutions to the problem of optimum stratification.

Dalenius and Gurney (1951) suggested boundary points, that give constant value for $W_h \sigma_h$ for all strata, to be a good approximation to optimum stratification under Neyman allocation.

Mahalanobis (1952) proposed the rule $W_h\mu_h$ = constant, to be a good approximation to the problem. In other words the boundary points are chosen in such a way that the strata totals are equal. With this rule, the expected contribution of each stratum to the total aggregate value of "X" is equal for all strata. Stevens (1952) also studied the advantages of equal number of units per stratum.

Hansen, Hurwitz and Madow (1953) demonstrated that equalisation of strata totals would give a good approximation to optimum stratification, if the different strata have a constant coefficient of variation and that in such situations, this approximation was almost similar to the approximation of Dalenius and Gurney (1951).

Aoyama (1954) derived the equi-distant stratification rule, X_{h+1}-X_h = constant by applying the mean value theorem to the original equation

$$\frac{\sigma_h^2 + (x_h - \mu_h)^2}{\sigma_h} = \frac{\sigma_{h+1}^2 + (x_h - \mu_{h+1})^2}{\sigma_{h+1}}$$

He assumed that the variation of density to be small in each stratum

Dalenius and Hodges (1957) suggested equalisation of cumulative $\sqrt{f}$ in all strata to be a good approximation to the problem of optimum stratification under Neyman allocation. Dalenius (1957) gave an elaborate account of various methods of stratification. Dalenius and Hodges (1959) provided the optimality of cumulative $\sqrt{f}$

solution assuming rectangular distribution with in each stratum, which is valid at least when the number of strata is large. They also demonstrated cumulative $\sqrt{f}$ rule to be good for small number of strata as well, in the case of some continuous populations. They also had given an iterative procedure to determine the optimum strata boundaries.

Durbin (1959) proposed a rule to form the strata by taking equal areas under a frequency distribution that is half way between the original and a rectangular distribution. The boundaries were to be obtained by taking equal intervals of $\gamma(x) + f(x)$, where, $\gamma(x) = F(x) / (x_L - x_O)$ and $F(x)$ is the cumulative distribution function associated with $f(x)$. This was arrived at as a first order correction to the rule of equalisation of strata ranges.

Ekman (1959), assuming regularity of the density function over finite range, suggested equalisation of the product of frequency and range with in each stratum

$$\text{i.e,} \quad W_h (Y_h - Y_{h-1}) = C, \qquad \text{for } h = 1,2,3,\text{---------}L$$

Cochran (1961) compared the approximate solutions of Dalenius and Hodges (1959), Mahalanobis (1952), Ekman (1959) and Durbin (1959) with the exact solutions for eight frequency distributions. He concluded that the approximations suggested by Dalenius and Hodges and Ekman performed consistently well under both optimum and equal allocations.

Aoyama (1963) suggested equal strata ranges to be a good approximation to optimum stratification, assuming rectangular distribution with in strata. Sethi (1963,1964) observed that the approximations of Mahalanobis (1952) and Dalenius and Gurney (1951) did not yield good strata boundaries in the case of certain continuous distributions. He reported that cumulative $\sqrt{f}$ method of determining strata boundaries lead to more efficient stratification than equalisation of strata totals or ranges for a truncated gamma distribution.

Raj (1964) studied the principle of equi partition of Mahalanobis (1952) on four populations having continuous distributions viz.,

1. $(2/\pi)^{1/2} \exp(-y^2), y \geq 0$

2. exp (-y), $y \geq 0$
3. y.exp(-y), $y \geq 0$
4. 2 (1-y), $0 \leq y \leq 1$

He found that the rule produced poor boundaries when the number of strata was large. Raj (1964) verified different approximate rules for optimum stratification for two, three and four strata cases for the standard normal distribution truncated at the positive side.

Kish (1965) criticised the rule of equal strata ranges on the ground that the extreme strata would be having very low frequencies resulting in too unequal strata. He was also critical of the rule of equi strata sizes as the extreme strata might be having very long strata ranges. He argued that the actual solution to the problem of optimum stratification has to be a compromise between these two rules. He also concluded that stratification that equalises value of $W_h\sigma_h$ would be satisfactory and that such stratification would lead to Neyman allocation. Thus he established that optimal stratification under Neyman allocation would lead to equal allocation.

Taga (1971) proved that optimum stratification on empirical distributions converged to that of the distributions of continuous univariate random variables.

Unnithan (1978) demonstrated the use of a modified Newton's procedure of function minimum for the problem of optimum stratification. He also showed that the available procedures could lead only to a locally optimum solution and that these might fail in situations when more than one minimum of the variance function exists.

Unnithan and Nair (1995) proposed an iterative procedure to arrive at global optimum of a function and used the same to determine the demarcation points of strata that minimise the sampling variance of the estimator of the population mean from a stratified simple random sample under Neyman and proportional allocations. They considered the study character itself as the stratification variable.

2.2 Stratification based on auxiliary variable

All the work referred to in section 2.1 are related to optimum stratification on the study character itself. But in practice we never have the required information on the study variable and the information is only available for some highly correlated auxiliary variable. Such literatures are reviewed in the following lines.

Hess, Sethi and Balakrishnan (1966) compared four methods of optimum stratifications applied to populations with strong positive skewness and concluded that Ekman's rule out performed the other methods

Murthy (1967) compared stratification rules of equal strata sizes, equal strata totals and equal values of the product of strata range and sizes (R_hW_h) under proportionate, Neyman, proportionate to strata total and proportionate to R_hW_h allocations using actual populations. He concluded that approximation of strata by equalising R_hW_h under the allocation of sample size in proportion to R_hW_h would provide good approximation to the problem of optimum stratification and allocation.

Taga (1967) suggested a procedure to arrive at optimum boundary points of stratification using apriori information, when the stratification is done on a separate variable or variables under proportional allocation.

Serfling (1968) suggested the use of cumulative $\sqrt{f}$ rule of Dalenius and Hodges (1957) for the construction of strata on the auxiliary variable when the regression of the study variable 'y' on the stratification variable 'x' is linear and the correlation is positive and very high.

Singh and Sukhatme (1969) derived equations to be satisfied by the boundary points to achieve the minimum variance under proportional and Neyman allocations, when stratification is done on a concomitant variable and the regression of the study character on the concomitant characters as well as the conditional variance of the study character on concomitant character are known. They also suggested approximate solutions, as the equations are not directly solvable.

Singh (1971a) studied the determination of optimum strata boundaries, optimum number of strata and sample allocation when the regression of the study variable 'y' on the stratification variable 'x' is linear and the correlation is positive and very high. He proposed cumulative $\sqrt[3]{f}$ rule for the construction of strata. According to him, cumulative $\sqrt[3]{f}$ is not a generalisation of cumulative $\sqrt{f}$ rule, as it does not reduce to the latter rule even when the correlation between 'y' and 'x' is one. Singh (1971b) gave a detailed discussion on the problem of optimum stratification for proportional allocation.

Cochran (1972) made a detailed account of the various approximations available for optimum allocation. Garg and Murty (1972) compared different methods of construction of strata on the basis of an auxiliary character in livestock surveys for estimating the number of milch cows and the milk production from the sample survey conducted in Kerala state during 1964-65. The stratification variable used is the number of milch cows as per the livestock census, 1961. They concluded that equal aggregate out put method and Ekman's methods were more efficient than cumulative $\sqrt{f}$ method.

Singh and Sukhatme (1972) gave certain asymptotic properties of the approximately optimum strata boundaries that could be obtained by following the method suggested by Ekman (1959), when stratification is done on an auxiliary variable whose regression with the study variable is known.

Singh and Sukhatme (1973) obtained rules for optimum stratification for regression and ratio type estimators of the population mean when the concomitant variable is linearly related to the study character.

Singh and Dev (1975) considered various approximations including those suggested by Singh and Sukhatme (1969) for equal allocation and recommended a cumulative $\sqrt{f}$ rule for approximate optimum strata boundaries, when stratification is done on a concomitant variable.

Singh (1975a) proposed an extension of the cumulative $\sqrt{f}$ rule for constructing strata on the basis of auxiliary variable. He also made an empirical investigation into the efficiency of the proposed rule with those suggested by Serfling (1968) and Singh

(1971) when the regression of the study variable (y) on the auxiliary variable (x) is of the form y = x. Singh (1975b) suggested approximately optimum stratification under proportional allocation.

Anderson, Kish and Cornell (1976) determined optimum strata boundaries under Neyman allocation, when stratification is done on a concomitant variable for varying levels of the correlation coefficient between the estimation variable (y) and the stratification variable (x), using a bivariate normal population. These boundaries were compared to those obtained through cumulative $\sqrt{f}$ and cumulative $\sqrt[3]{f}$ methods.

Thomsen (1976) found equal intervals of cumulative cube root 'f' to give approximately optimum strata boundaries when the stratification is carried out on an auxiliary variable which compared favorably, in certain situations, with those determined by the cumulative root 'f' rule.

Prakash and Bokil (1978) compared different methods of construction of strata based on an auxiliary character to determine the strata boundaries along with the number of strata in cultivated fodder surveys to determine the area under fodder crops in 1972-73. He used the area under fodder crops in 1970-71 as the stratification variable and concluded that cumulative $\sqrt[3]{f}$ rule of stratifying the population gave the smallest variance compared to all the other methods.

Kish and Anderson (1978) proposed another approach based on two stratification variates. They used cumulative $\sqrt{f}$ rule on one variate to generate a set of strata and then repeated the procedure on the other.

Sethumadhavi and Sukhatme (1979) did an investigation of stratification for a population with high skewness. They considered the method of constructing strata, the sample allocation, the number of strata and the optimum sample size for estimating the number of trees and total area under fruits using the information on area under fresh fruits as the stratification variable in Mahasu of Himachal Pradesh. Comparisons were made among four types of allocation in combination with the corresponding stratifications. They concluded that the rules proposed by Dalenius and Hodges (1957), Ekman (1959), Mahalanobis (1952) and Durbin (1959) gave nearly identical results. However, beyond two strata, Ekman's rule excelled followed by that of Mahalanobis (1952) and Dalenius and Hodges (1957).

Bhatnagar and Banerjee (1982) dealt with the problem of optimum stratification in sample surveys for estimating the total yield of cereal crops by using the information on the holding size of cultivators selected. They concluded by choosing the cumulative root 'f' rule for four strata as the best procedure.

Bose *et al.* (1982) made an attempt to obtain the optimum points of stratification on the basis of an auxiliary variable by arranging the value of the auxiliary variable in descending order and then dividing the set into as many groups required. The performance of the proposed rule of stratification has been examined empirically on a variety of natural populations. They found that the method is superior to the existing methods if the distribution of the auxiliary variable is positively skewed.

Poddar and Rustogi (1983) dealt with the problem of finding optimum points of stratification in sample surveys for estimating area under HYV of cereal crops in Ambala district of Haryana state. They used the information on the holding size of the cultivators selected for the area estimation enquiry in the district as the stratification variable. After comparing different methods of stratification, they concluded that cumulative $\sqrt[3]{f}$ rule was optimally best for seasonal crops.

Sukhatme *et al.* (1984) provided a good discussion on different methods of finding optimum strata boundaries. Iachan (1985) reviewed the problem of finding stratum boundaries that minimised the variance of the survey estimates. The practical situation when stratification is on auxiliary variables was examined and applied to bottom trawl surveys of shellfish.

Unnithan (1995) extended the iterative procedure of Unnithan and Nair (1995) to arrive at the optimum boundary points of stratification under Neyman and proportional allocation, when stratification is done on a separate stratification variable, which is closely related to the study characteristic.

Mehta *et al.* (1996) derived equations to be satisfied by the boundary points of strata in terms of an auxiliary variable, that minimises the sampling variances of ratio, regression and product estimators of population mean of the study character, under allocation of sample size in proportion to strata totals. They also proposed approximate solution, as the equations cannot be easily solved. The lower limits to the variance of the estimators of the population mean under the above three methods are also obtained and thus the behavior of variance with increasing number of strata can be identified.

2.3 Multivariate approach

Multivariate sampling received early attention by Snedecor and King (1942). But the emphasis was on allocation problems in already defined strata. Neyman (1934) suggested that if the survey variates are highly correlated, allocation for a specific variate would yield reasonable allocation for the other variates as well. If not, allocation proportional to stratum size would be the best.

An attempt to extend Dalenius (1950) theory to a bivariate population (x,y) was made by Ghosh (1963). He studied two-way stratification based on two characters under study and discussed optimum boundary points, which minimised the generalised variance of unbiased linear estimator.

Stratification which minimises the generalised variance of estimates of the means of more than one character under study based on an auxiliary character was investigated by Gupta and Seth (1979) for proportionate allocation assuming the number of strata. They concluded that if the regression of each of the study character on the auxiliary character is of the same form, the points of stratification based on the auxiliary character yielded estimates which had the smallest concentration ellipsoid among all the points of stratification based on the auxiliary character.

Sadasivan and Agarwal (1974) derived the optimum points of stratification with two variables under study by extending the exact solutions by Dalenius (1950) to two variate case with the same two as the stratification variables. Minimising the generalised variances, they got a set of equations giving the optimum points of stratification for varying and constant correlation between the two variables.

Rao (1979) studied the problem of optimum stratification for a multivariate population when the simultaneous estimation of several independent characters are involved. He also described an approximate rule for quick determination of strata boundaries in multivariate populations.

Jat *et al.* (1979) studied the problem of stratification in multivariate population. They made an empirical comparison among the different methods for determining the strata boundaries for multivariate population, when two stratification variables are involved using two real populations to estimate two study characters by using the generalised variance of sample means for the study characters.

Pla (1992) described the use of the first principal component as stratification variable to determine optimum strata boundaries in multivariate sampling to improve precision in estimating the mean vector.

MATERIALS AND METHODS

MATERIALS AND METHODS

The materials for the present study of comparison of optimum strata boundaries determined by various methods consisted of 1025 trees of Forestro variety of cocoa (*Theobroma cacao*.L) planted in December 1988 at the Cadbury-KAU Co-operative Cocoa Research Project (CCRP), College of Horticulture, Kerala Agricultural University, Thrissur. The farm has a latitude of 10° 31' North, a longitude of 76° 17' East and is at an elevation of 25m above MSL. The soil type was lateritic loam. The spacing of the trees was 3m x 3m. The manurial and cultural practices were done as per the package of practice recommendation of the Kerala Agricultural University.

Observations considered in this investigation were copied from the field records of the project. The following are the observations taken.

- Pod yield

The numbers of pods harvested from each plant for four consecutive years *viz.*, 1995-96, 1996-97, 1997-98 and 1998-99 were utilized.

- Plant height

Plant height recorded for each tree during the year 1991-92 was utilized.

- Canopy spread

Arithmetic mean of canopy spread of each tree in cm in the North- South and East- West directions recorded in the year 1991-92 was worked out.

- Trunk girth

Trunk girth measured at a height of 15 cm from ground during the year 1991-92 for each tree was the fourth character considered.

3.1 Methods

Three approximate methods making use of the frequency distribution of the population along with the iterative method to obtain the global minimum of an objective function utilised for stratification was attempted in this investigation. The stratification by each of these methods was carried out on a number of stratification variables. Their efficiency was assessed by estimating the sampling variance of the estimator of population mean of the study character from the stratified random sample under Neyman allocation.

3.1.1 Study variable

Study variable in the present investigation was pod yield during 1998-99.

3.1.2 Stratification variable

Stratification was carried out on an auxiliary variable / variables that are believed to have a strong association with future yield. Such a procedure would lead to efficient stratification, provided stratification variables and study variable are closely related. The following variables (auxiliary) were used for stratification in the present investigation.

3.1.2.1 Average yield of yester years

Present yield of any perennial crop is considered as a strong indication of future yield. Many tree crops have a biennial nature of bearing. Considering these facts, the arithmetic mean of yield for each of the 1025 cocoa trees during 1995-96 and 1996-97 was one of the stratification variables considered.

3.1.2.2 Canopy spread

Canopy is a character that measures the photosynthetic activity of the plant and hence must have a strong influence on the yield. Hence the arithmetic mean of canopy spread in cm in North- South and East- West directions was a character in which stratification was attempted.

3.1.2.3 $H \times G^2$, { H= Height of the plant & G= Trunk girth }

This derived variable has been taken as an index of seedling vigor in cocoa and has been widely used for selection of seedlings world over. It gives an indication of the tree volume and hence is a good measure of tree health and hence must have a strong influence on the yield. Thus HxG^2 was considered as another stratification variable.

3.1.2.4 The first principal component.

A plant is characterized by a number of characters and to study it, individual characters in isolation can give only a partial picture. In other words, if important characters are considered simultaneously, we get comprehensive information on the tree. Thus the first principal component of the following variates was considered as a stratification variable to treat the multivariate situation as a univariate situation.

- Plant height
- Plant girth
- Canopy spread
- HxG^2, and
- Average yield of yester years

3.1.2.5 Regression estimate

The first principal component is that linear combination of variates which retains maximum variability contained in the component characters. This derived variable need not be the best linear function of those component characters that influences the future yield, the study characters under consideration. The multiple linear regression estimator of the future yield is the best linear predictor of the study characters. Thus the multiple linear regression estimator of the study characters, with the following predictor variables was considered as a stratification variable in the present investigation.

- Average yield of yester years

- Canopy spread
- $H x G^2$: and
- First principal component

3.1.3 Number of strata

Two strata, three strata, four strata and five strata situations were considered for stratification in the case of various stratification variables and different methods of stratification in the present investigation.

3.1.4 Type of allocation

The solution to the problem of optimum stratification under Neyman allocation of sample size to different strata had been receiving attention due to the complicated nature of the equation to be satisfied by the strata boundaries. Moreover it is also optimum allocation in a sense. Therefore for efforts to determine strata boundaries in this investigation, Neyman allocation of sample size was assumed.

3.1.5 Sample size

A sample of size 200 trees was chosen for estimation of sampling variance in all cases.

3.1.6 Formation of frequency distributions

Frequency distributions of the cocoa trees on the basis of each of the stratification variables was prepared and are given in the appendices II, III, IV, V and VI. The methods of quick determination of demarcation points of strata were based on these frequency distributions.

3.1.7 Methods of stratification

Since the theoretical solutions to the equations to be satisfied by optimal strata boundaries under Neyman allocation are not available, a number of approximations have been suggested. Three of these that were adjudged best by various workers (Prakash and Bokil, 1978: Bhatnagar and Banerjee, 1982: Poddar and Rustogi, 1983, etc.) in different empirical situations have been chosen for the present investigation and were compared. In addition to these three, an iterative procedure for obtaining the points of stratification that achieves the global optimum of the sampling variance function of the estimator of the population mean from a stratified simple random sampling (Unnithan, 1995) was also included.

3.1.7.1 Method 1 – Equalisation of cumulative root 'f'

Dalenius and Hodges (1957) proposed formation of strata by equalising the cumulative of $\sqrt{f}$ where 'f' is the frequency function. In the frequency distribution, the square roots of frequencies were obtained in each class and the cumulative distribution of these square root values was prepared for each of the stratification variable. From this table of cumulative distribution, the strata boundaries for two to five strata situations are determined as detailed in the section 3.1.8.

3.1.7.2 Method 2- Equalisation of cumulative cube root 'f'

Equalisation of cumulative $\sqrt[3]{f}$ was proposed by Singh (1971a) for the formation of strata. In this method, the procedure is similar to that described in the section 3.1.7.1 for cumulative root 'f' method except that cube root values of frequencies were obtained in each class instead of square root values for each of the stratification variables.

3.1.7.3 Method-3 Ekman's method

In this method, the product of the strata size and width were equalised for each stratum. Accordingly within each class of the frequency distribution corresponding to each stratification variable, the product of the class width and the frequency was obtained in each class and the cumulative values of these quantities were obtained just as the cumulative frequency distribution. Once these values are obtained, the approximate strata boundaries were obtained as described in the section 3.1.8.

3.1.7.4 Iterative method

This method dealt with a procedure to arrive at a set of boundary points of stratification that achieves global minimum of the variance function. In this method, we consider all boundary points of stratification with a restriction that any stratum is continuous. From among all stratifications we choose that which give the minimum value for the function F (x). According to Unnithan (1995), when stratification is carried out on a concomitant variable using Neyman allocation, the sampling variance of the population mean of the study characteristic in terms of the variance of the stratification variable, is given by

$$F(x) = \sum_{h=1}^{L} Wh\left[\sigma_{hx}{}^{2} + \sigma^{2}{}_{x}\frac{(1-p)}{p}\right]^{1/2}$$

where, Wh = size of the h^{th} stratum.

$\sigma^2{}_{hx}$ = variance of the concomitant variable in the h^{th} stratum

$\sigma^2{}_{x}$ = variance of the concomitant variable

p = proportion of variation explained by the concomitant variable

Divide the interval (A, B), the range of the variable on which stratification is assumed to be carried out, into a large number of small intervals such that no interval thus formed shall have more than one minimum of F(x). Initially search for the minimum of F(x) is made from among a finite number of selected possible demarcation points, namely $x_1, x_2, x_3, \ldots\ldots, x_{N-1}$. This can be done in the case of 'L' strata as follows.

1. Fix the values of the first L-2 boundary points at $x_1, x_2, x_3, \ldots\ldots, x_{L-2}$ and then the minimum of F(x) is searched by evaluating F(x) at all possible values of x_{L-1}. i.e., find

$$\underset{x_{L-1} \le x_{L-1} \le x_{N-1}}{\text{Min}} \left[F(x)/x_1, x_2, \ldots\ldots\ldots x_{L-2} = x_{L-2}\right] \quad \text{------------------------(1)}$$

2. Advance the values of x_{L-2} by one point. i.e., let x_{L-2} be assumed x_{L-1} and search for the minimum of F(x) is made among all possible values of x_{L-1}. i.e, find

$$\underset{x_{L} \le x_{L-1} \le x_{N-1}}{\text{Min}} \left[F(x)/x_1, x_2, \ldots\ldots\ldots x_{L-2} = x_{L-1}\right] \quad \text{------------------------(2)}$$

3. Similarly evaluate minimum of F(x) by all possible values of x_{L-2} as in steps 1 and 2 and choose that value for x_{L-2} and x_{L-1} for which F(x) is minimum. That is for which F(x) arrived at by (1) and (2) is minimum.

4. Choice for x_{L-2} and x_{L-1} for fixed values of $x_1, x_2, x_3, \ldots\ldots, x_{L-3}$ at $x_1, x_2, x_3, \ldots\, x_{L-3}$ have been made so as to make F(x) a minimum. Next, the value of x_{L-3} is advanced by one point and the minimum of F(x) among the minimum values of F(x) arrived at as in step 3. The choice of x_{L-3}, x_{L-2} and x_{L-1} can be made by choosing the minimum of the minimum F(x) values corresponding to different values of x_{L-3}.

5. The procedure is continued till we arrive at the choice of all the L-1 stratification points which make F(x) a minimum. i.e., we can arrive at the minimum of F(x) as

$$\underset{x}{\mathrm{Min}}F(x) = \underset{x_1}{\mathrm{Min}}\,\underset{x_2}{\mathrm{Min}}\ldots\ldots\ldots\underset{x_{L-1}}{\mathrm{Min}}\left[F(x)/x_1 = x_{1,}\ldots\ldots\ldots x_{L-2} = x_{L-2}\right]$$

Thus we arrive at the best stratification points to be used as the first approximation. Further improvement in this solution can be done as follows.

The two smaller intervals on either sides of each stratification point obtained may be collapsed and the same may be divided into much smaller intervals. Then all combinations of stratification points, by allocating these smaller intervals to one of the two adjoining strata in the case of each stratification point with the restriction that intervals allotted to a stratum form a continuous interval, every time evaluating F(x) may be considered. Choose that configuration which gives the minimum value of F(x). This process may be continued until the required accuracy for the boundary points is achieved.

The computer programme in BASIC used for determining the strata boundaries in this method is given in the appendix-1.

3.1.8 Determination of strata boundaries

The equation used for the determination of strata boundaries for two to five strata for approximate methods 3.1.7.1, 3.1.7.2, and 3.1.7.3 is given by

$$X_i = l + \left[\frac{iS}{L} - m\right]\frac{c}{W_c}$$

where, 'l' is the lower limit of the interval containing the boundary point between i^{th} and $i+1^{st}$ strata for i = 1,2,3,.....L-1

'S' is the cumulative total of $\sqrt{f}$ or $\sqrt[3]{f}$ or $f_h (X_h - X_{h-1})$ as the case may be

'm' is the corresponding cumulative total upto the interval containing the strata boundary

'W_c' is the contribution of the class containing the strata boundary towards 'S'

'c' is the class interval of the class containing the strata boundary: and

'L' is the no. of strata, L = 1,2,3,4,5.

3.1.9 Comparison of the methods

Comparison of these methods is done by comparison of estimates of the sampling variance of the estimator of the population mean of the study character. The estimate of variance under Neyman allocation is given by

$$V(\overline{y_{st}}) = \frac{\left(\sum_{i=1}^{L} WiSi\right)^2}{n} - \frac{\left(\sum_{i=1}^{L} WiSi^2\right)}{N}$$

where 'Wi' the i^{th} stratum size (Wi= Ni/ N)

'Si' the square root of mean square for the i^{th} stratum

'n' the total sample size

'N' the population size

'L' the number of strata

RESULTS

RESULTS

The data were analysed using the methods described in the chapter 'Materials and Methods'. Results obtained by different stratification rules based on different stratification variables are presented below.

4.1 Stratification on average yield of yester years.

The frequency distribution of the 1025 cocoa trees by average yield of yester years was prepared (appendix-II) and the strata boundaries were formed based on this distribution for different number of strata. The strata boundaries determined by the cumulative root 'f', cumulative cube root 'f', Ekman's and the iterative methods for two to five strata and the corresponding estimate of the variance of the sample mean are presented in the Table 1.

4.1.1 Cumulative root 'f' method

First stratum consisted of trees having an average yield of yester years upto 21 pods and those trees having the yield greater than 21 pods belonged to second stratum in the case of two strata by the method of cumulative root 'f'. Estimate of the variance of the population mean under Neyman allocation was 1.5787

In the case of three strata, those trees having the average yield upto 13 pods belonged to first stratum, those having yield above 13 pods, but upto 29 pods comprised the second stratum. Trees that yield above 29 pods formed the third stratum. The estimated variance for the mean was 1.5356

In the case of four strata, those trees having the average yield upto 10 pods belonged to first stratum and those above 10 pods, but upto 20 pods comprised the second stratum. Trees that yield 21 to 34 pods formed the third stratum and those above 34 pods formed the fourth stratum. The estimate of variance was 1.4761

In the case of five strata the boundary points were 8 for first stratum, 16 for second stratum, 25 for third stratum, 37 pods for fourth stratum. Trees with average

Table 1. Strata boundaries and variance estimate for varying number of strata using the average yield of yester years as the stratification variable

No. of Strata	Strata boundaries for different rules				Estimate of variance of the estimator			
	Cuml. √f	Cuml. ^{3}√f	Ekman	Iterative	Cuml. √f	Cuml. ^{3}√f	Ekman	Iterative
2	21	22	29	29	1.5787	1.6616	1.6695	1.6695
3	13	15	18	17	1.5356	1.5382	1.5403	1.5446
	29	31	44	41.5				
4	10	11	13	13	1.4761	1.4784	1.4809	1.4830
	20	22	30	29				
	34	36	49	50.5				
5	8	9	10	11.5	1.4851	1.4937	1.4921	1.4677
	16	18	21	25.5				
	25	27	41	44				
	37	39	52	76				

yield above 37 pods belong to fifth stratum. The variance of the sample mean was estimated to be 1.4851 and was higher than that in the case of four strata.

4.1.2 Cumulative cube root 'f' method

For cumulative cube root 'f' rule, the first stratum consisted of trees having an average yield upto 22 pods and those trees having the yield greater than 22 pods belonged to the second stratum in the case of two strata. The variance of the mean under Neyman allocation was estimated to be 1.6616

In the case of three strata, those trees having the average yield upto 15 pods belonged to first stratum and those between 15 and 31 pods comprised the second stratum. Trees that yield above 31 pods formed the third stratum. The variance of the mean under Neyman allocation was estimated to be 1.5382

In the case of four strata, those trees having the average yield upto 11 pods formed the first stratum and those between 11 and 22 pods comprised the second stratum. Trees that yielded 22 to 36 pods formed the third stratum and those above 36 pods formed the fourth stratum. The estimate of the variance for the mean decreased and reached at 1.4784

In the case of five strata the boundary points were 9 for first stratum, 18 for second stratum, 27 for third stratum, 39 pods for fourth stratum. Trees with average yield above 39 pods belong to fifth stratum. As in the method 1 the variance estimate showed a slight increase to reach at 1.4937.

4.1.3 Ekman's method

For Ekman's rule, the first stratum consisted of trees having an average yield upto 29 pods and those trees having the yield greater than 29 pods belonged to second stratum in the case of two strata. The variance of the mean under Neyman allocation was given by 1.6695

In the case of three strata, those trees having the average yield upto 18 pods belonged to first stratum and those above 18, but upto 44 pods comprised the second stratum. Trees that yield above 44 pods formed the third stratum. The estimate of the variance for the mean was found to be 1.5403

In the case of four strata, those trees having the average yield upto 13 pods belonged to first stratum, those trees having yield above 13 pods and less than or equal to 30 pods comprised the second stratum. Trees that yielded 30 to 49 pods formed the third stratum and those above 49 pods formed the fourth stratum. The variance estimate for the mean still decreased and reached at 1.4809

In the case of five strata the boundary points were 10 for first stratum, 21 for second stratum, 41 for third stratum, 52 pods for fourth stratum. Trees with average yield above 52 pods belong to fifth stratum. In this case also, like above two methods, the variance shows a slight increase and reached at 1.4921.

4.1.4 Iterative method

For iterative method, the first stratum consisted of trees having an average yield upto 29 pods and those trees having the yield greater than 29 pods belonged to second stratum in the case of two strata. The variance of the mean under Neyman allocation was given by 1.6695

In the case of three strata, those trees having the average yield upto 17 pods belonged to first stratum and those between 17 and 41.5 pods comprised the second stratum. Trees that yield above 41.5 pods formed the third stratum. The variance of the mean under Neyman allocation was found to be 1.5446

In the case of four strata, those trees having the average yield upto 13 pods belonged to first stratum and those between 13 and 29 pods comprised the second stratum. Trees that yield 29 to 50.5 pods formed the third stratum and those above 50.5 pods formed the fourth stratum. The estimate of the variance for the mean was 1.4830

In the case of five strata the boundary points were 11.5 pods for first stratum, 25.5 pods for second stratum, 44 pods for third stratum, 76 pods for fourth stratum. Trees with average yield above 76 pods belonged to fifth stratum.The boundary points given was included in the lower strata. The estimate of the variance for the sample mean was given by 1.4677

4.2 Stratification on canopy spread.

The frequency distribution of the 1025 cocoa trees by canopy spread was prepared (appendix- III). Strata boundaries were determined by the cumulative root 'f', cumulative cube root 'f' and Ekman's methods using the frequency distribution for two to five strata. Demarcation points of strata were also obtained for two to five strata by the iterative method. The boundary points so determined by the four methods along with the corresponding estimates of sampling variance of the estimator of population mean of the study character are provided in Table 2.

4.2.1 Cumulative root 'f' method

The first strata consisted of trees having canopy spread upto 219cms and those trees having the spread above 219cms belonged to second stratum in the case of two strata by the cumulative root 'f' rule. The estimate of the variance of the sample mean under Neyman allocation was 2.1462

In the case of three strata, those trees having the canopy spread upto 168cms belonged to first stratum and those between 168 and 269cms comprised the second stratum. Trees that had canopy spread above 269cms formed the third stratum. The variance estimate was 2.1094

In the case of four strata, those trees having the canopy spread upto 140cms were included in the first stratum and those between 140 and 219cms in the second stratum, those between 219 to 295cms in the third stratum and those above 295cms in the fourth stratum. The estimate of the variance of the population mean was 2.0964

Table 2. Strata boundaries and variance estimate for varying number of strata using the canopy spread as the stratification variable

No. of Strata	Strata boundaries for different rules				Estimate of variance of the estimator			
	Cuml. $\sqrt{f}$	Cuml. $\sqrt[3]{f}$	Ekman	Iterative	Cuml. $\sqrt{f}$	Cuml. $\sqrt[3]{f}$	Ekman	Iterative
2	219	221	95	167.75	2.1462	2.1417	2.0308	2.1119
3	168	169	18	87.75	2.1094	2.1121	2.0222	2.0370
	269	272	198	230.25				
4	140	140	13	65.25	2.0964	2.0994	2.0060	2.0178
	219	221	95	177.75				
	295	299	243	270.25				
5	120	122	11	52.5	2.0563	2.0543	1.9988	2.0157
	189	190	22	143.5				
	249	252	163	219				
	307	312	269	294				

In the case of five strata the boundary points were 120cms for first stratum, 189cms for second stratum, 249cms for third stratum and 307 for fourth stratum. Trees with canopy spread above 307cms belonged to the fifth stratum. The boundary points given were included in the lower strata. In this case the estimated variance of the estimator was found to be 2.0563

4.2.2 Cumulative cube root 'f' method

When strata were formed by cumulative cube root 'f' rule, the first stratum consisted of trees having the canopy spread upto 221cms and those trees having the spread above 221 cms belonged to second stratum in the case of two strata. The estimate of variance of the mean under Neyman allocation was 2.1417

In the case of three strata, those trees having the canopy spread upto 169cms belonged to first stratum and those between 169 and 272cms comprised the second stratum. Trees that had canopy spread above 272cms formed the third stratum. The estimated variance of the sample mean was 2.1121

In the case of four strata, those trees having the canopy spread upto 140cms were allocated the first stratum and those between 140 and 221cms to the second stratum. Trees with canopy spread 221 to 299cms formed the third stratum and those above 299cms formed the fourth stratum. The estimated variance of the estimator was 2.0994

In the case of five strata the boundary points were 122cms for first stratum, 190cms for second stratum, 252cms for third stratum and 312 for fourth stratum. Trees with canopy spread above 312cms belong to fifth stratum. The boundary points given were included in the lower strata. The sampling variance of the estimator was estimated to be 2.0543

4.2.3 Ekman's method

By adopting Ekman's rule, the first stratum was found to consist of trees having canopy spread upto 95cms and those trees having spread above 95 cms belonged to second stratum in the case of two strata. The estimate of the variance was given by 2.0308

In the case of three strata, trees having canopy spread upto 18cms were grouped to first stratum and those between 18cms and 198cms were grouped to the second stratum. Trees that had canopy spread above 198cms formed the third stratum. The estimate of the variance of the sample mean was found to be 2.0222

In the case of four strata, trees having canopy spread upto 13cms were in the first stratum and those between 13 and 95cms were in the second stratum. Trees with canopy spread 95 to 243cms formed the third stratum and those above 243cms formed the fourth stratum. The variance of the estimator was estimated to be 2.0060

In the case of five strata the boundary points were 11cms for first stratum, 22cms for second stratum, 163cms for third stratum and 269cms for fourth stratum. Trees with canopy spread above 269cms belonged to the fifth stratum. The boundary points given were included in the lower strata. Sampling variance of the estimator was estimated as 1.9988

4.2.4 Iterative method

Using the iterative method, the first stratum was found to consist of trees having canopy spread upto 167.75cms and trees having spread greater than 167.75cms constituted the second stratum in the case of two strata. The variance of the mean under Neyman allocation was 2.1119

In the case of three strata, trees having canopy spread upto 87.75cms belonged to first stratum and those between 87.75 and 230.25cms comprised the second stratum. Trees that had spread above 230.25cms formed the third stratum. The variance of the mean under Neyman allocation was estimated to be 2.0370

In the case of four strata, trees having canopy spread upto 65.25cms formed the first stratum and those between 65.25 and 177.75cms comprised the second stratum. Trees that had spread from 177.75 to 270.25cms constituted the third stratum and those above 270.25 formed the fourth stratum. The estimate of the variance for the mean was 2.0178

In the case of five strata the boundary points were 52.5 for first stratum, 143.5 for second stratum, 219 for third stratum, 294 for fourth stratum. Trees with canopy spread above 294cms belonged to fifth stratum. The estimate of the variance for the sample mean was 2.0157

4.3 Stratification on HxG^2

The frequency distribution of the 1025 cocoa trees according to the value of HxG^2 was prepared (appendix- IV) and the strata boundaries were determined by the cumulative root 'f', cumulative cube root 'f', Ekman's and the iterative methods used for two to five strata. Sampling variances of the were also estimated and are provided in Table 3 along with the strata boundaries.

4.3.1 Cumulative root 'f' method

When strata were constructed by cumulative root 'f' rule, the first stratum consisted of trees having the value of HxG^2 upto 44379 and the trees having a value of above 44379 belonged to the second stratum in the case of two strata. The estimated variance of the estimator in this case was 2.0761

In the case of three strata, trees having value of HxG^2 upto 26585 belonged to first stratum and those between 26585 and 85427 comprised the second stratum. Trees that had HxG^2 value above 85427 formed the third stratum. The estimated variance was found to be 2.0124

In the case of four strata, trees having value of HxG^2 upto 18502 formed first stratum and those between 18502 and 44379 comprised the second stratum. Trees with value of HxG^2 ranging from 44379 to 115634 formed the third stratum and those above 115634 formed the fourth stratum. The estimated variance for the mean was 2.0105

Table 3. Strata boundaries and variance estimate for varying number of strata using the value of H x G^2 as the stratification variable

No. of Strata	Strata boundaries for different rules				Estimate of variance of the estimator			
	Cuml. $\sqrt{f}$	Cuml. $\sqrt[3]{f}$	Ekman	Iterative	Cuml. $\sqrt{f}$	Cuml. $\sqrt[3]{f}$	Ekman	Iterative
2	44379	39950	184850	97480	2.0761	2.0535	2.1854	2.1370
3	26585	24909	119825	67510	2.0124	2.0301	2.1337	2.1162
	85427	71332	304566	153900				
4	18502	17595	95554	45570	2.0105	2.0188	2.1130	2.0579
	44379	39949	184850	106010				
	115634	106574	353425	188500				
5	13126	13740	83514	45600	2.0098	2.0068	2.0973	2.0520
	35502	26290	138139	105100				
	64194	48853	265480	183500				
	133930	128798	382740	283500				

In the case of five strata the boundary points were 13126 for first stratum, 35502 for second stratum, 64194 for third stratum and 133930 for fourth stratum. Trees with the value of HxG^2 above 133930 belonged to fifth stratum. Estimate of the sampling variance was found to be 2.0098

4.3.2 Cumulative cube root 'f' method

When strata were constructed by cumulative cube root 'f' method, the following boundaries for two, three, four and five strata situations was obtained. The first stratum consisted of trees having value of HxG^2 upto 39950 and trees having the value above 39950 belonged to the second stratum in the case of two strata. The estimate of the variance was given by 2.0535

In the case of three strata, those trees having value of HxG^2 upto 24909 belonged to first stratum and those between 24909 and 71332 comprised the second stratum. Trees that had the value above 71332 formed the third stratum. The estimated sampling variance of the estimator was 2.0301

In the case of four strata, trees having value of HxG^2 upto 17595 belonged to first stratum and those between 17595 and 39949 comprised the second stratum. Trees with value of HxG^2 ranging from 39949 to 106574 formed the third stratum and those above 106574 formed the fourth stratum. The estimated variance was found to be 2.0188

In the case of five strata the boundary points were 13740 for first stratum, 26290 for second stratum, 48853 for third stratum and 128798 for fourth stratum. Trees with value of HxG^2 above 128798 belonged to fifth stratum. Sampling variance of the estimator was estimated to be 2.0068

4.3.3 Ekman's method

Boundary points of stratification for two to five strata situation was determined by the Ekman's rule and the details are given below. The first stratum consisted of trees having value of HxG^2 upto 184850 and those having value above

184850 belonged to second stratum in the case of two strata. The estimate of variance of the sample mean is given by 2.1854

In the case of three strata, trees having value of HxG^2 upto 119825 formed the first stratum and those between 119825 and 304566 comprised the second stratum. Trees that had HxG^2 value above 304566 formed the third stratum. The estimated sampling variance was found to be 2.1337

In the case of four strata, trees having value of HxG^2 upto 95554 belonged to first stratum and those between 95554 and 184850 comprised the second stratum. Trees with value of HxG^2 ranging from 184850 to 353425 formed the third stratum and those above 353425 formed the fourth stratum. The estimated variance for the mean in this case was 2.1130

In the case of five strata the boundary points were 83514 for first stratum, 138139 for second stratum, 265480 for third stratum and 382740 for fourth stratum. Trees with the value of HxG^2 above 382740 belonged to fifth stratum. Sampling variance was estimated to be 2.0973

4.3.4 Iterative Method

When the iterative method was adopted for determination of strata boundaries for two to five strata situations, the results were obtained as follows. The first stratum consisted of trees having value of HxG^2 upto 97480 and those having value above 97480 belonged to second stratum in the case of two strata. The estimate of the variance of sample mean was given by 2.1370

In the case of three strata, those trees having value of HxG^2 upto 67510 constituted the first stratum and those between 67510 and 153900 comprised the second stratum. Trees that had HxG^2 value above 153900 formed the third stratum. The estimated variance was found to be 2.1162

In the case of four strata, those trees having value of HxG^2 upto 45570 belonged to first stratum and those between 45570 and 106010 comprised the second stratum. Trees with value of HxG^2 ranging from 106010 to 188500 constituted the third stratum and those above 188500 formed the fourth stratum. The estimated variance for the mean was found to be 2.0579

In the case of five strata the boundary points were 45600 for first stratum, 105100 for second stratum, 183500 for third stratum and 283500 for fourth stratum. Trees with value of HxG^2 above 283500 formed fifth stratum. Sampling variance was estimated at 2.0520

4.4 Stratification on the first principal component

Stratification was attempted based on the first principal component of the characters that was considered as stratification variables individually. The correlation-half matrix of the six variates mentioned is chapter 3 is given in the Table 4a and the principal components were calculated. The latent roots, percentage variance and the cumulative variance of the first five principal components are given in the Table 4b. Since the first principal component explains maximum variance (67.58%), it can be used as a stratification variable.

The frequency distribution of the 1025 cocoa trees according to the first principal component was prepared (appendix-V) and the strata boundaries was arrived at based on this distribution for different number of strata by the cumulative root 'f', cumulative cube root 'f', Ekman's and the iterative methods for two to five strata situations. Estimates of the variance of the sample mean were also obtained and were presented in Table 5.

4.4.1 Cumulative root 'f' method

Strata boundaries arrived at by the cumulative root 'f' method for two to five strata situations, considering the first principal component as the stratification variable, are given below. The first stratum consisted of trees having value of the first principal component upto 33525 and those having value above 33525 belonged to second stratum in the case of two strata. The sampling variance in this case was estimated to be 2.1438

Table 4a. Correlation half matrix of the six variates

	Pl. Height	Pl.Girth	Av.Spread	HxG^2	Av.Yld.
Pl.Height	1.0000	0.7830	0.7772	0.6102	0.4004
Pl.Girth		1.0000	0.7357	0.9006	0.3425
Av.Spread			1.0000	0.5593	0.3487
HxG^2				1.0000	0.2504
Av.Yld.					1.0000

Table 4b. Latent roots and variance explained by principal components

PRIN#	Latent Roots	Percentage Variance	Cumulative Variance
Prin. 1	3.379	67.589	67.589
Prin. 2	0.841	16.811	84.400
Prin. 3	0.506	10.116	94.516
Prin. 4	0.216	4.330	98.845
Prin. 5	0.058	1.155	100.000

Table 5. Strata boundaries and variance estimate for varying number of strata using the value of first principal component as the stratification variable

No. of Strata	Strata boundaries for different rules				Estimate of variance of the estimator			
	Cuml. $\sqrt{f}$	Cuml. $\sqrt[3]{f}$	Ekman	Iterative	Cuml. $\sqrt{f}$	Cuml. $\sqrt[3]{f}$	Ekman	Iterative
2	33525	34610	96963	44860	2.1438	2.1431	2.1230	2.1674
3	21257	22343	77527	43510	2.0834	2.0912	2.1120	2.1378
	47772	48087	116057	128400				
4	15242	16376	71673	30630	2.0362	2.0448	2.0089	2.1227
	33525	34610	96963	68950				
	55486	55544	105051	153910				
5	11450	12491	57422	19650	2.0117	2.0366	2.1026	2.0069
	26280	27178	83800	46300				
	41510	42210	108697	80600				
	60214	62036	130777	165000				

In the case of three strata, those trees having value of the first principal component upto 21257 belonged to first stratum and those between 21257 and 47772 comprised the second stratum. Trees that had value above 47772 formed the third stratum. The estimate of the variance of the mean was found to be 2.0834

In the case of four strata, trees having value of the first principal component upto 15242 belonged to first stratum and those between 15242 and 33525 comprised the second stratum. Trees with value ranging from 33525 to 55486 formed third stratum and those above 55486 formed the fourth stratum. The estimated variance for the mean was 2.0362

In the case of five strata the boundary points were 11450 for first stratum, 26280 for second stratum, 41510 for third stratum and 60214 for fourth stratum. Trees with values above 60214 belonged to fifth stratum. The variance of the estimator was estimated at 2.0117

4.4.2 Cumulative cube root 'f' method

When cumulative cube root 'f' method was used to determine the strata boundaries for two to five strata situations, the following results were obtained. The first stratum consisted of trees having value of the first principal component upto 34610 and those having the value above 34610 belonged to second stratum in the case of two strata. The estimate of variance was given by 2.1431

In the case of three strata, those trees having value of the first principal component upto 22343 belonged to first stratum and those between 22343 and 48087 comprised the second stratum. Trees that had values above 48087 formed the third stratum. The estimated variance of the mean was found to be 2.0912

In the case of four strata, those trees having value of the first principal component upto 16376 belonged to first stratum and those between 16376 and 34610 comprised the second stratum. Trees with value ranging from 34610 to 55544 formed the third stratum and those above 55544 formed the fourth stratum. The variance estimate was determined to be 2.0448

In the case of five strata the boundary points are 12491 for first stratum, 27178 for second stratum, 42210 for third stratum and 62036 for fourth stratum. Trees with values above 62036 belonged to fifth stratum. The variance of the estimator was estimated at 2.0366

4.4.3 Ekman's method

When Ekman's rule was applied for stratification on the first principal component for two to five strata cases, the boundary points arrived at are as follows. The first stratum consisted of trees having value of the first principal component upto 96963 and those having value above 96963 belonged to second stratum in the case of two strata. The value of the estimate of variance was given by 2.1230

In the case of three strata, those trees having value of the first principal component upto 77527 belonged to first stratum and those between 77527 and 116057 comprised the second stratum. Trees that had value above 116057 formed the third stratum. The estimated variance for the mean was found to be 2.1120

In the case of four strata, those trees having the value of the first principal component upto 71673 constituted the first stratum and those between 71673 and 96963 comprised the second stratum. Trees with value ranging from 96963 to 105051 formed the third stratum and those above 105051 formed the fourth stratum. The variance estimate for the mean was determined at 2.0089

In the case of five strata the boundary points were 57422 for first stratum, 83800 for second stratum, 108697 for third stratum and 130777 for fourth stratum. Trees with the value above 130777 belonged to fifth stratum. The variance of the estimator was estimated to be 2.1026

4.4.4 Iterative method

Iterative method was also used on the first principal component as stratification variable and the demarcation points of strata were determined as follows. The first stratum consisted of trees having value of the first principal component upto 44860 and those having value above 44860 belonged to second stratum in the case of two strata. The value of the estimate of variance was given by 2.1674

In the case of three strata, those trees having value of the first principal component upto 43510 belonged to first stratum and those between 43510 and 128400 comprised the second stratum. Trees that had value above 128400 formed the third stratum. The sampling variance was estimated to be 2.1378

In the case of four strata, those trees having value of the first principal component upto 30630 belonged to first stratum and those between 30630 and 68950 comprised the second stratum. Trees with value ranging from 68950 to 153910 formed the third stratum and those above 153910 formed the fourth stratum. The variance estimate for the mean was estimated as 2.1227

In the case of five strata the boundary points were 19650 for first stratum, 46300 for second stratum, 80600 for third stratum and 165000 for fourth stratum. Trees with value above 165000 belonged to fifth stratum. The sampling variance was estimated at 2.0069

4.5 Stratification on the regression estimate

The frequency distribution of the 1025 cocoa trees by the values of regression estimates was prepared (appendix- VI) and the strata boundaries were determined based on this distribution for different number of strata by the cumulative root 'f', cumulative cube root 'f' and Ekman's methods. Iterative method was also used to determine the strata boundaries for two to five strata situations. Estimates of the sampling variance of the estimator of the population mean was also obtained and are provided in Table 6 along with the strata boundaries.

4.5.1 Cumulative root 'f' method

By cumulative root 'f' rule, the first stratum consisted of trees having regression estimates upto 25 pods and trees having pods above 25 belonged to

Table 6. Strata boundaries and variance estimate for varying number of strata using the regression estimate as the stratification variable

No. of Strata	Strata boundaries for different rules				Estimate of variance of the estimator			
	Cuml. $\sqrt{f}$	Cuml. $\sqrt[3]{f}$	Ekman	Iterative	Cuml. $\sqrt{f}$	Cuml. $\sqrt[3]{f}$	Ekman	Iterative
2	25	25	42	31	1.6602	1.6602	1.8208	1.7014
3	20	20	29	29	1.5292	1.5888	1.6151	1.6630
	30	31	47	82				
4	17	18	24	21.5	1.4637	1.4986	1.5412	1.4968
	25	25	42	34				
	34	35	50	54				
5	16	16	21	16	1.4610	1.4559	1.5132	1.4958
	22	22	37	26				
	28	29	45	38				
	35	36	53	58.5				

second stratum in the case of two strata and the variance estimate was given by 1.6602

In the case of three strata, trees having regression estimates upto 20 pods belonged to first stratum and those between 20 and 30 pods comprised the second stratum. Trees that had pod yield above 30 formed the third stratum. The estimate of the variance for the estimator was found to be 1.5292

In the case of four strata, trees having regression estimates upto 17 pods were grouped into the first stratum and those between 17 and 25 pods were grouped into the second stratum. Trees with values ranging from 25 to 34 pods formed the third stratum and those above 34 formed the fourth stratum. The estimated variance was 1.4637

In the case of five strata the boundary points were 16 for first stratum, 22 for second stratum, 28 for third stratum and 35 for fourth stratum. Trees with the estimated yield of above 35 pods formed the fifth stratum. The variance was estimated as1.4610

4.5.2 Cumulative cube root 'f' method

When cumulative cube root 'f' rule was adopted, the first stratum consisted of trees having regression estimates upto 25 pods and those trees having the value above 25 belonged to second stratum in the case of two strata. Here the strata configuration was exactly same as that obtained by the cumulative root 'f' rule. The estimate of variance was given by 1.6602

In the case of three strata, trees having regression estimates upto 20 pods belonged to first stratum and those between 20 and 31 pods comprised the second stratum. Trees that had the estimates of above 31 pods formed the third stratum. The estimated variance of the estimator was 1.5888

In the case of four strata, trees having regression estimates upto 18 pods belonged to first stratum and those between 18 and 25 pods comprised the second stratum. Trees with estimates ranging from 25 to 35 pods formed the third stratum and those above 35 pods constituted the fourth stratum. The variance estimate obtained was 1.4986

In the case of five strata the boundary points were 16 for first stratum, 22 for second stratum, 29 for third stratum and 36 for fourth stratum. Trees with estimated yield of above 36 pods belonged to fifth stratum. The variance was estimated to be 1.4559

4.5.3 Ekman's method

Demarcation points of strata was determined by the Ekman's rule for two to five strata situations taking the regression estimates as the stratification variables and the following results were obtained. The first stratum consisted of trees having regression estimates upto 42 pods and those having the value above 42 pods belonged to second stratum in the case of two strata. The value of the estimate of variance was given by 1.8208

In the case of three strata, trees having regression estimates upto 29 pods belonged to first stratum and those between 29 and 47 comprised the second stratum. Trees that had estimated yield of above 47 pods formed the third stratum. The estimated variance for the estimator was found to be 1.6151

In the case of four strata, those trees having regression estimates upto 24 pods were included in the first stratum and those between 24 and 42 pods were included in the second stratum. Trees with the regression estimates ranging from 42 to 50 pods formed the third stratum and those above 50 pods formed the fourth stratum. The variance estimate was 1.5412

In the case of five strata the boundary points were 21 for first stratum, 37 for second stratum, 45 for third stratum and 53 for fourth stratum. Trees with the estimated yield of above 53 pods belonged to fifth stratum. The variance was estimated at 1.5132

4.5.4 Iterative method

Iterative procedure of determination of strata boundaries was adopted with the regression estimates as stratification variable in the case of two to five strata situations and the strata configuration arrived at are given below. The first stratum consist of trees having regression estimates upto 31 pods and those having values above 31 pods belonged to second stratum in the case of two strata. The value of the estimate of variance was given by 1.7014

In the case of three strata, those trees having regression estimates upto 29 pods belonged to first stratum and those between 29 and 82 pods comprised the second stratum. Trees that had estimated yield of above 82 pods formed the third stratum. The estimated variance for the estimator was found to be 1.6630

In the case of four strata, trees having regression estimates upto 21.5 pods belonged to first stratum and those between 21.5 and 34 pods comprised the second stratum. Trees with regression estimates ranging from 34 to 54 pods formed the third stratum and those above 54 pods formed the fourth stratum. The variance estimate was obtained as 1.4968

In the case of five strata the boundary points were 16 for first stratum, 26 for second stratum, 38 for third stratum and 58.5 for fourth stratum. Trees with estimated yield of above 58.5 pods belonged to fifth stratum. The sampling variance of the estimator was estimated at 1.4958

Discussion

DISCUSSION

Most attempts for the solution of optimum stratification have treated the study character itself as the stratification variable for simplicity. But stratification can be done only on concomitant variable / variables, which is closely related to the study character and information on which are available before the survey. The attempt herein is to compare different methods of stratification, when it is done on a separate stratification variable by assessing their efficiency through the estimate of the sampling variance of the estimator of population mean.

Three stratification rules that were proposed as approximations to the solution of the problem of optimum stratification, *viz.*, cumulative root 'f' rule, cumulative cube root 'f' rule and Ekman's rule and adjudged good by many research workers were considered in this investigation. These methods were originally proposed for the situation of treating study character itself as the stratification variable. The same rules were used for separate stratification variables in this attempt. Along with these rules, an iterative solution proposed recently was also made use of.

Optimum strata boundaries arrived at for two, three, four and five strata by different methods based on five different stratification variables, viz, average yield of yester years, canopy spread, HxG^2 value, the first principal component and the regression estimate, are given in Tables 1,2,3,5 and 6. These are discussed in this chapter.

5.1 Comparison of different stratification methods

Relative change in the sampling variance of the estimator of population mean with respect to the different stratification methods was used for comparison. The sampling variance obtained for stratification done by cumulative $\sqrt{f}$ rule was taken as the reference method and the sampling variances in the other methods were expressed as ratio to the reference method and are presented in Table 7.

Table.7 Relative changes in sampling variance in different methods using stratification variables

No. of Strata	Cuml. √f	Cuml. ³√f	Ekman	Iterative
Average yield of yester years				
2	1	1.0525	1.0575	1.0575
3	1	1.0017	1.0031	1.0059
4	1	1.0016	1.0033	1.0047
5	1	1.0058	1.0047	0.9882
Canopy spread				
2	1	0.9979	0.9462	0.9840
3	1	1.0013	0.9587	0 9657
4	1	1.0014	0.9569	0.9625
5	1	0.9990	0.9720	0.9803
HxG^2				
2	1	0.9891	1.0526	1.0293
3	1	1.0088	1.0603	1.0516
4	1	1.0043	1.0510	1.0236
5	1	0.9985	1.0435	1.0209
First principal component				
2	1	0.9997	0.9903	1.0110
3	1	1.0037	1.0137	1.0261
4	1	1.0042	0.9866	1.0425
5	1	1.0124	1.0452	0.9976
Regression estimate				
2	1	1	1.0967	1.0248
3	1	1.0390	1.0562	1.0875
4	1	1.0238	1.0529	1.0226
5	1	0.9965	1.0357	1.0238

5.1.1 Average yield of yester years as stratification variable

In the case of stratification based on average yield of yester years, the cumulative root 'f' method was found to be superior since it gave a comparatively low estimate of the variance of the estimator followed by the cumulative cube root 'f' rule for the case of two, three and four strata. Ekman's method and iterative method gave equal variances in two strata case. In the case of three and four strata case, Ekman's method performed better than iterative method.

But in the case of five strata, the iterative method excelled giving a smaller estimate of variance of the estimator compared to all other methods.

5.1.2 Canopy spread as stratification variable

In the case of two strata, three strata and four strata and five strata, Ekman's method came superior in the case of stratification based on the canopy spread. Iterative method performed better than the two cumulative methods. Cumulative cube root 'f' method edged out the cumulative root 'f' method in the case of two and five strata and in the case of three and four strata, cumulative root 'f' method gave slightly lower variance than the cumulative cube root 'f' rule.

5.1.3 HxG^2 as stratification variable

When stratification was done on the value of HxG^2, we can see a mixed performance of the different methods. In the case of two strata and five strata, cumulative cube root 'f' method produced comparatively low estimate of variance followed by the cumulative root 'f method. Iterative method gave a slightly lower variance than the Ekman's method in these cases.

In the case of three strata and four strata, cumulative root 'f method came superior followed by the cumulative cube root 'f' method. Just as the two and five strata case, the iterative procedure gave a slightly lower variance than the Ekman's method in these cases.

5.1.4 First principal component as stratification variable

In the case of stratification on the first principal component also a mixed performance of different stratification rules in different strata situations can be observed. Ekman's method resulted in least variance in the case of two strata. Cumulative root 'f' and cumulative cube root 'f' methods gave almost identical results in two and four strata situations where the cumulative cube root 'f' method gave a slightly lower variance of the estimator. Iterative procedure resulted in largest variance.

In the case of three strata, the cumulative root 'f' method came superior compared to all other methods followed by the cumulative cube root 'f' method. Ekman's method gave a comparatively lower variance than the iterative method .

In the case of four strata, the Ekman's method was found to be superior compared to all other methods giving a smaller variance estimate than all other methods as in the case of two strata followed by the two cumulative methods. Iterative method gave a reasonably higher variance than all the other methods in this situation also just as the above two methods.

But in the case of five strata, the iterative procedure resulted in a comparatively low sampling variance for the estimator. The cumulative root 'f' and cube root methods followed the iterative procedure. Ekman's method gave a slightly higher variance.

5.1.5 Regression estimate as stratification variable

In the case of stratification on regression estimate, the cumulative root 'f' method seemed to be superior in the case of two, three and four strata situations. Cumulative cube root 'f' rule gave identical result with that of the cumulative root 'f' method in the case of two strata and followed the same pattern in the case of three strata. In the case of two strata, the iterative method follows the two cumulative methods and in the case of three strata, the Ekman's procedure followed the two methods. In the case of four strata, the iterative procedure came second followed by the cumulative cube root 'f' and the Ekman's method.

In the case of five strata, the cumulative cube root 'f' rule is found to be superior compared to all other methods and is followed by the cumulative root 'f' method. Iterative method gave a better variance estimate compared to the Ekman's method in this situation.

5.2 Comparison of different stratification variables

Relative changes in the sampling variance of the estimator of population mean with respect to the different stratification variables were used for comparison. The sampling variance obtained by using the average yield of yester years as stratification variable was taken as the reference variable. Ratios of sampling variances under each stratification variables to that under average yield of yester years are presented in Table 8.

When strata was constructed by cumulative root 'f' method using regression estimate as the stratification variable, sampling variance of the estimator was found to be least compared to other stratification variables except in the case of two strata where it was higher than that of the average yield of yester years.

While cumulative cube root 'f' method was used for determination of strata boundaries, the regression estimate was found to be the best stratification variable for two and five strata situations followed by the average yield of yester years and the average yield of yester years found as best stratification variable for three and four strata situations followed by the regression estimates.

For Ekman's and iterative methods of stratification, average yield of yester years was found to be the best stratification variable closely followed by regression estimate irrespective of the number of strata.

Table 8. Relative changes in sampling variance for different stratification variables under various methods

No. of Strata	Average yield	Canopy spread	HxG^2	1stprincipal component	Regression estimate
Cumulative √f method					
2	1	1.3959	1.3151	1.3980	1.0516
3	1	1.3739	1.3105	1.3507	0.9958
4	1	1.4202	1.3620	1.3794	0.9916
5	1	1.3846	1.3533	1.3456	0.9838
Cumulative ^{3}√f method					
2	1	1.2889	1.2359	1.2898	0.9992
3	1	1.3731	1.3198	1.3595	1.0329
4	1	1.3592	1.3655	1.3831	1.0136
5	1	1.3753	1.3435	1.3635	0.9746
Ekman's method					
2	1	1.2164	1.3090	1.2716	1.0906
3	1	1.3129	1.3852	1.3711	1.0486
4	1	1.3546	1.4268	1.3565	1.0407
5	1	1.3396	1.4056	1.4092	1.0141
Iterative method					
2	1	1.2650	1.2800	1.2982	1.0191
3	1	1.3188	1.3700	1.3840	1.0766
4	1	1.3606	1.3877	1.4314	1.0093
5	1	1.3734	1.3981	1.3674	1.0191

Average yield of yester years can be adjudged as the best stratification variable in general. Regression estimates can also be recommended as stratification variable for estimating yield of cocoa. These variables can also be used to calibrate cocoa trees in order to form blocks for conducting planned experiments.

5.3 Effect of number of strata in sampling variance

Relative changes in sampling variance of the estimator of population mean with respect to different number of strata were also used for their comparison. The sampling variance obtained for two strata situation was taken as the reference number and the sampling variances in the case of other numbers of strata were expressed as ratio to the reference sampling variance and are presented in Table 9.

Sampling variance decreased with the increase in number of strata when iterative procedure was used for determination of strata boundaries. In some of the other cases, there was an increase in sampling variance when number of strata increased from four to five, which is contrary to expectation. The rate of decrease in sampling variance in the case of iterative procedure seems to suggest to be more suitable when number of strata is large. Generally, the sampling variance decreased with increase in number of strata. When average yield of yester years was taken as the stratification variable, there was an increase in sampling variance when number of strata increased from 4 to 5. A similar increase was also noticed in the sampling variance when number of strata was increased from 4 to 5 when Ekman's procedure was used to determine the strata boundaries with the first principal component as the stratification variable. It is worth noting that in the case of iterative procedure sampling variance decreased with increase in number of strata in case of all stratification variables.

5.4 Conclusion

Cumulative root 'f' method was found to have resulted in lower sampling variance in general and the difference between root 'f' and cube root 'f' rules was very marginal. The iterative procedure fared well in the sense of lower sampling variance when the number of strata was five. Ekman's procedure fared well only

Table.9 Relative changes in sampling variance in different number of strata using stratification variables

No. of Strata	Cuml. √f	Cuml. $^3\sqrt{f}$	Ekman	Iterative
Average yield of yester years				
2	1	1	1	1
3	0.9727	0.9257	0.9226	0.9252
4	0.9350	0.8897	0.8870	0.8882
5	0.9407	0.8989	0.8946	0.8791
Canopy spread				
2	1	1	1	1
3	0.9828	0.9862	0.9958	0.9645
4	0.9768	0.9802	0.9877	0.9554
5	0.9581	0.9592	0.9842	0.9544
HxG²				
2	1	1	1	1
3	0.9693	0.9886	0.9763	0.9903
4	0.9684	0.9831	0.9669	0.9630
5	0.9680	0.9773	0.9597	0.9602
First principal component				
2	1	1	1	1
3	0.9718	0.9758	0.9948	0.9863
4	0.9498	0.9541	0.9463	0.9794
5	0.9384	0.9502	0.9904	0.9259
Regression estimate				
2	1	1	1	1
3	0.9211	0.9570	0.8870	0.9794
4	0.8816	0.9027	0.8464	0.8797
5	0.8800	0.8769	0.8311	0.8791

when the canopy spread was used as the stratification variable. The inconsistency would be because of the instability in the yield of the trees considered in the investigation. The study character was the yielding for the year 1998-99. It could be inferred that when number of strata is large, the iterative procedure could be recommended and for lower number of strata, cumulative root 'f' method seems to be good enough.

Regarding stratification variable, average yield of yester years was found to be the best and the regression estimate also could be recommended. So these variables can also be used as covariate to perform analysis of covariance (ANCOVA) for experiments in cocoa. Also the suggested stratification variables and the procedure of stratification can be successfully used for formation of blocks when the experiments are planned for the crop.

SUMMARY

SUMMARY

A comparison among some of the promising methods proposed by various workers for determination of optimum strata boundaries was attempted in this investigation. The methods are compared by estimating the sampling variance of the estimator of the population mean under Neyman allocation of sample size to different strata in two to five strata situations when stratification was done on a number of stratification variables. A sample size of 200 was used in all methods and stratification variables of the investigation.

The investigation was made on data on 1025 trees of Forestro variety of cocoa (*Theobroma cacao*.L) planted in December 1988 at the Cadbury-KAU Co-operative Cocoa Research Project (CCRP), College of Horticulture, Kerala Agricultural University, Thrissur. The observations were yield of four consecutive years from 1995-96, canopy spread in both the North- South and East- West directions, plant height and the trunk girth. The yield for 1998-99 was taken as the study character.

The stratification variables considered were the average yield of yester years, canopy spread, HxG^2, where H and G are respectively the height and girth of the tree, the first principal component derived from plant height, plant girth, canopy spread, HxG^2 and the average yield of yester years and the regression estimate of the study character with the average yield of yester years, canopy spread, HxG^2 and the first principal component derived earlier as the predictor variables.

The following were the different stratification procedures tried.

5.1 Cumulative root 'f method

This method was proposed by Dalenius and Hodges (1957) and the strata are formed by equalising the cumulative of the square root 'f', where 'f' is the frequency in various classes in the frequency distribution of the 1025 trees by the concerned stratification variable.

5.2 Cumulative cube root 'f' method

Singh (1971a) proposed this method for finding strata boundaries. The procedure is similar to the above method except that the cube root of the frequency function was used instead of the square root values.

5.3 Ekman's method

The Ekman's method of determining demarcation points of strata consisted of equalising the product of the strata size and the strata width for each stratum. With in each class of frequency distribution corresponding to each stratification variable, the product of class width and the frequency was obtained and the cumulative of these values are prepared.

5.4 Iterative procedure

This method dealt with a procedure to arrive at the set of boundary points of stratification that achieves global minimum of the variance function. In this method, we consider all boundary points of stratification with a restriction that any stratum is continuous. From among all stratifications we choose that which gives the minimum value for the variance function. According to Unnithan (1995), when stratification is carried out on a concomitant variable using Neyman allocation, the sampling variance of the population mean of the study characteristic in terms of the variance of the stratification variable, is given by

$$F(x) = \sum_{h=1}^{L} Wh\left[\sigma_{hx}^{\ 2} + \sigma^2_{\ x}\,\frac{(1-p)}{p}\right]^{1/2}$$

where, Wh = size of the h^{th} stratum.

$\sigma^2_{\ hx}$ = variance of the concomitant variable in the h^{th} stratum

$\sigma^2_{\ x}$ = variance of the concomitant variable

p = proportion of variation in the study character explained by the concomitant variable

The results obtained from the investigation are summarized below.

1. Among the different stratification methods, the performances of the methods were found to be varying. The performance of the stratification rules depended on the number of strata and the stratification variable. For smaller number of strata, the cumulative root and cube root 'f' methods were found to be good in the case of stratification based on the average yield, regression estimate, HxG^2 and on the first principal component. Ekman's procedure was found to be superior in the case of stratification based on canopy spread. But as the number of strata increases, the iterative procedure was found to be consistently better in the performance in almost all cases.

2. Among the different stratification variables considered, average yield of yester years was found to be the best followed by the regression estimate in the sense of resulting in smaller sampling variance for the estimator. Stratification based on HxG^2 and on the first principal component were found to be inferior.

REFERENCES

REFERENCES

Anderson, D. W., Kish, L. and Cornell, R.G. 1976. Quantifying gains from stratification for optimum and approximately optimum strata using a bivariate normal model. *J. Am. stat. Assoc.* **55**: 887-892

Aoyama, H. 1954. A study of the stratified random sampling. *Ann. Math. Stat.* **6**:1-36

Aoyama, H. 1963. Stratified random sampling with optimum allocation for multivariate populations. *Ann. Inst. stat. Math.* **14**: 251-258

Balasubramnian, P.P. 1999. *Cashew and Cocoa- Vast Untapped Potential.* Kerala Society for Horticultural Science. p,44-48

Bhatnagar, K. C. and Banerjee, A.K. 1982. Optimum points of stratification for estimating yield of cereal crops. *J. Indian Soc. agric. Stat.* **34**(3): 100

Bose, R. K., Banerjee, A. K. and Dey ,A . 1982. A simple rule for stratification. *J. Indian Soc. agric. Stat.* **34**(3): 102

Cochran, W.G. 1961. Comparison of methods of determining strata boundaries. *Bull. int. stat. Math.* **38** : 345-358

Cochran, W.G. 1972. *Sampling Techniques.* Wiley Eastern Private Ltd., New Delhi, p.89-146

Dalenius, T. 1950. The problem of optimum stratification. *Skandinavik Aktuarietidskrift* **33**: 203-213

Dalenius, T. 1957. *Sampling in Sweden.* Almqvist and Wiksell, Stockholm, p.159-168

Dalenius, T. and Gurney, M. 1951. The problem of optimum stratification II. *Skandinavik Aktuarietidskrift* **34**: 133-148

Dalenius, T. and Hodges, J. L. 1957. The choice of stratification points. *Skandinavik Aktuarietidskrift* **3-4**: 198-203

Dalenius, T and Hodges, J.L.1959. Minimum variance stratification. *J. Am. stat. Assoc.* **54**: 88-101

Durbin, J. 1959. Review of sampling in Sweden. *J. Roy. stat. Soc.*(A), **122**: 246-248

Ekman, G .1959. An approximation useful in univariate stratification. *Ann. Math. Stat.* **30**: 219-229

Garg, R. C. and Murty, V.V. R. 1972. Efficiency of stratification in livestock surveys. *J. Indian Soc. agric. Stat.* **24**(2): 55-64

Ghosh, S. P. 1963. Optimum stratification with two characters. *Ann. Math. Stat.* **34**: 866-872

Gupta, P. C. and Seth, G. R. 1979. On stratification in sampling investigation involving more than one character. *J. Indian Soc. agric. Stat.***31**(2): 1-13

Hansen, M. H., Hurwitz, W. N. and Madow, W. G. 1953. *Sample Survey Methods and Theory*. John Wiley and sons, New York, p.

Hess, I., Sethi, V.K. and Balakrishnan, T.R. 1966. Stratification. A practical investigation. *J. Am. stat. Assoc.* **61**: 74-90

Iachan, R. 1985. Optimum stratum boundaries for shellfish surveys. *Biometrics* **41**(4): 1053-1062

Jat, P.R., Rao, D.V.S.and Jain, J.P. 1979. Comparison of methods for determining stratum boundaries in multivariate populations. *J. Indian Soc. agric. Stat.* **31**(3): 76

Kish, L. 1965. *Survey Sampling.* John Wiley and sons, Inc. New York, p.100-110

Kish, L. and Anderson, D. W. 1978. Multivariate and multipurpose stratification. *J. Am. stat. Assoc.* **73**: 24-34

Mahalanobis, P.C. 1952. Some aspects of Design of sample surveys. *Sankhya* **12**: 1-7

Mehta, S. K., Singh, R. and Kishore, L. 1996. On optimum stratification for allocation proportional to strata totals. *J. Indian stat. Assoc.* **34**: 9-19

Murthy, M. N. 1967. *Sampling Theory and Methods.* Statistical Publishing Society, Calcutta, p.233-287

Nampoothiri, K.V.K. and Balasimha, D. 1999. Areca and cocoa- Viable Alternative. *The Hindu Survey of Indian Agriculture*, p.97-100

Neyman, J. 1934. On the two different aspects of the representative method : The method of stratified sampling and the method of purposive selection. *J. Roy. stat. Soc.* **97**: 558-606

Pla, L. 1992. Determining stratum boundaries with multivariate real data. *Biometrics* **47**(4): 1409-1422

Poddar, J. and Rustogi, V. S. 1983. Optimum points of stratification for estimating area under HYV of cereal crops. *J. Indian Soc. agric. Stat.* **35**(2): 165

Prakash, A. and Bokil, S.D. 1978. Stratification and estimation procedure in cultivated fodder surveys. *J. Indian Soc. agric. Stat* .**30**(2): 52-58

Rao, D. V. S. 1979. Stratification in multivariate population-II. *J. Indian Soc. agric. Stat.* **17**: 75

Raj, D.1964. On formation of strata of equal size. *J. Am. stat. Assoc.* **59**: 481-486

Raj, D. 1968. *Sampling Theory* . Tata McGraw Hill Publishing Company Ltd. New Delhi, p. 61-80

Sadasivan, G. and Agarwal, R. 1974. Optimum points of stratification in multi dimensional populations. *J. Indian Soc. agric. Stat.* **24**(1): 146

Serfling, R.J. 1968. Approximately Optimum Stratification . *J. Am. stat. Assoc.* **63**: 1298-1309

Sethi, V.K. 1963. A note on optimum stratification of populations for estimating the population means. *Aust. J. Stat.* **5**: 20-33

Sethi, V.K. 1964. Contributions to stratified sampling and some related problems. Ph.D. thesis, Institute of Social Sciences, Agra, New Delhi.

Sethumadhavi, R. and Sukhatme, B. V. 1979. Stratification in surveys on fruit crops. *J. Indian Soc. agric. Stat.* **31**(1): 63-76

Singh, R. 1971a. Approximately optimum stratification on the auxiliary variable. *J. Am. Stat. Assoc.* **66**: 829-833

Singh, R. 1971b. Determination of optimum strata boundaries. *Indian Soc. agric. Stat.* **23**(2):115-123

Singh, R. 1975a. An alternative method of stratification on the auxiliary variable. *Sankhya,* series. C. **37**:100-108

Singh, R. 1975b. On optimum stratification for proportional allocation. *Sankhya,* series. C. **37**:109-115

Singh, R. and Dev, P. 1975. Optimum stratification for equal allocation. *Ann. Inst. stat. Math.* **27**: 273-280

Singh, R. and Sukhatme, B. V. 1969. Optimum Stratification. *Ann. Inst. stat. Math.* **21**: 515-525

Singh, R. and Sukhatme, B. V. 1972. A note on optimum stratification. *J. Indian Soc. agric. Stat.* **24**(2): 91-98

Singh, R. and Sukhatme, B. V. 1973. Optimum Stratification. *Ann. Inst. Stat. Math.* **25**: 515-528

Snedecor, G. and King, A.J. 1942. Recent developments in sampling for agricultural statistics. *J. Am. stat. Assoc.* **37**: 95-102

Stevens, W.L. 1952. Sampling with same number on each stratum. *Biometrika* **39**: 414-416

Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Asok, C. 1984. *Sampling theory of Surveys with Applications.* Indian Society of Agricultural Statistics, New Delhi, p. 80-135

Taga, Y .1967. On optimum stratification for the objective variable based on concomitant variables using prior information. *Ann. Inst. stat. Math.* **19**: 101-129

Taga, Y. 1971. On covergence of optimum stratifications for empiric distribution function in univariate case. *Ann. Inst. stat. Math.* **23**: 355-363

Thomsen, I. 1976. A comparison of approximately optimal stratification given proportional allocation with other methods of stratification and allocation. *Metrika.* **23**(1): 15-25

Unnithan, V.K.G. 1978: The minimum variance boundary points of stratification. *Sankhya*, series C. **40**: 60-72

Unnithan, V.K.G. 1995. Some contributions to the problem of optimum stratification. Ph.D thesis. Cochin University of Science and Technology, Cochin.

Unnithan, V.K.G. and Nair, N.U. 1995 Minimum variance stratification. *Commun. Stat.*, Simulation and computation. **24**(1). 275-284

APPENDICES

Appendix-I

Computer programme in BASIC used to obtain the strata boundaries for iterative procedure

```
REM  TO  OBTAIN  THE  MINIMAL SOLUTION  FOR OPTIMUM
REM STRATIFICATION  IN  FIVE  STRATA  SITUATION BASED
REM  ON A STRATIFICATION VARIABLE
DIM X (2000), S (10), SS (10), NN (10)
INPUT N, N1*
INPUT "DATA FILE NAME", D$
OPEN "I", #1, D$
FOR I = 1 TO N
INPUT #1, X (I)
NEXT I
FOR J  = 1 TO N1 - 4 STEP 1
FOR K = J + 1 TO N1 - 3 STEP 1
FOR L  = K + 1 TO N1 - 2 STEP 1
FOR M = L + 1 TO N1 - 1 STEP 1
FOR KK = 1 TO 5
NN (KK) = 0: S (KK) = 0: SS (KK) = 0
NEXT KK
FOR I = 1 TO N
IF X (I) >= J THEN 10
KK = 1
GOTO 30
10   IF X (I) >= K THEN 20
KK = 2
GOTO 30
20   IF X(I) >= L THEN 25
KK = 3
GOTO 30
25   IF X (I) >= M THEN 27
KK = 4
GOTO 30
```

```
27      KK = 5
30      NN (KK) = NN (KK) + 1
        S (KK) = S (KK) + X (I)
        SS (KK) = SS (KK) + X (I) * X (I)
        NEXT I
        VAR = 0
        FOR KK = 1 TO 5
        V = (SS (KK) / NN (KK) - (S (KK) / NN (KK)) ^ 2) + C**
        VAR = VAR + SQR (V) * (NN (KK) / N)
        NEXT KK
        SB = J: SC = K: SD = L: SE = M
        IF J > 1 THEN 50
        IF K > J + 1 THEN 50
        IF L > K + 1 THEN 50
        IF M > L + 1 THEN 50
        VAR1 = VAR
        SB1 = SB: SC1 = SC: SD1 = SD: SE1 = SE
50      IF VAR >= VAR1 THEN 60
        VAR1 = VAR
        SB1 = SB: SC1 = SC: SD1 = SD: SE1 = SE
60      NEXT M
        NEXT L
        NEXT K
        PRINT VAR1, SB1, SC1, SD1, SE1
        NEXT J
        PRINT "MINIMAL VARIANCE="; VAR1
        PRINT "STRATA BOUNDARY = "; SB1, SC1, SD1, SE1
        END
```

* Maximum value of the stratification variable (x) used

** The constant $\sigma^2_x(1-p)/p$

Appendix- II

Frequency distribution of 1025 trees based on the value of average yield of yester years

Class Interval	Frequency	Class Interval	Frequency
0-1	86	28-29	8
1-2	42	29-30	9
2-3	21	30-31	17
3-4	23	31-32	16
4-5	20	32-33	12
5-6	39	33-34	8
6-7	23	34-35	11
7-8	22	35-36	17
8-9	21	36-37	10
9-10	23	37-38	7
10-11	25	38-39	12
11-12	25	39-40	9
12-13	34	40-45	34
13-14	26	45-50	25
14-15	23	50-55	24
15-16	23	55-60	14
16-17	28	60-65	9
17-18	25	65-70	8
18-19	34	70-75	5
19-20	26	75-80	1
20-21	26	80-85	1
21-22	20	85-90	3
22-23	19	90-95	0
23-24	27	95-100	1
24-25	15	100-110	1
25-26	20	110-120	0
26-27	25	120-130	0
27-28	20	130-140	2

Appendix-III

Frequency distribution of 1025 trees based on the value of canopy spread

Class Interval	Frequency	Class Interval	Frequency
0-25	160	220-225	20
25-50	4	225-230	15
50-55	0	230-235	8
55-60	2	235-240	20
60-65	7	240-245	24
65-70	0	245-250	19
70-75	15	250-255	12
75-80	2	255-260	12
80-85	1	260-265	18
85-90	11	265-270	19
90-95	3	270-275	27
95-100	22	275-280	18
100-105	12	280-285	7
105-110	17	285-290	30
110-115	8	290-295	3
115-120	6	295-300	23
120-125	11	300-305	34
125-130	6	305-310	18
130-135	8	310-315	17
135-140	19	315-320	11
140-145	7	320-325	11
145-150	25	325-330	11
150-155	11	330-335	14
155-160	4	335-340	12
160-165	29	340-345	6
165-170	7	345-350	9
170-175	22	350-360	8
175-180	12	360-370	7
180-185	8	370-380	10
185-190	28	380-390	5
190-195	9	390-400	8
195-200	31	400-410	3
200-205	9	410-420	2
205-210	14	420-430	1
210-215	18	430-440	0
215-220	14	440-450	1

Appendix- IV

Frequency distribution of 1025 trees based on the value of HxG^2

Class Interval	Frequency	Class Interval	Frequency
0-1000	48	35000-36000	5
1000-2000	4	36000-37000	3
2000-3000	11	37000-38000	4
3000-4000	17	38000-39000	7
4000-5000	8	39000-40000	20
5000-6000	8	40000-41000	4
6000-7000	11	41000-42000	2
7000-8000	13	42000-43000	8
8000-9000	6	43000-44000	4
9000-10000	11	44000-45000	10
10000-11000	11	45000-46000	4
11000-12000	6	46000-47000	3
12000-13000	13	47000-48000	7
13000-14000	7	48000-49000	6
14000-15000	19	49000-50000	3
15000-16000	2	50000-60000	59
16000-17000	6	60000-70000	67
17000-18000	8	70000-80000	51
18000-19000	7	80000-90000	51
19000-20000	6	90000-100000	46
20000-21000	2	100000-110000	36
21000-22000	8	110000-120000	42
22000-23000	3	120000-130000	40
23000-24000	3	130000-140000	46
24000-25000	14	140000-150000	32
25000-26000	11	150000-160000	25
26000-27000	4	160000-170000	21
27000-28000	3	170000-180000	21
28000-29000	12	180000-190000	20
29000-30000	7	190000-200000	15
30000-31000	4	200000-210000	12
31000-32000	0	210000-220000	13
32000-33000	3	220000-250000	12
33000-34000	5	250000-500000	20
34000-35000	5		

Appendix- V

Frequency distribution of 1025 trees based on the value of the first principal component

Class Interval	Frequency	Class Interval	Frequency
0-1000	52	38000-39000	5
1000-2000	32	39000-40000	14
2000-3000	19	40000-41000	5
3000-4000	20	41000-42000	19
4000-5000	22	42000-43000	8
5000-6000	21	43000-44000	7
6000-7000	26	44000-45000	6
7000-8000	11	45000-46000	10
8000-9000	15	46000-47000	3
9000-10000	11	47000-48000	10
10000-11000	10	48000-49000	9
11000-12000	25	49000-50000	10
12000-13000	7	50000-51000	17
13000-14000	19	51000-52000	4
14000-15000	6	52000-53000	7
15000-16000	11	53000-54000	14
16000-17000	8	54000-55000	0
17000-18000	11	55000-56000	22
18000-19000	24	56000-57000	4
19000-20000	12	57000-58000	9
20000-21000	14	58000-59000	10
21000-22000	12	59000-60000	2
22000-23000	9	60000-61000	15
23000-24000	16	61000-62000	9
24000-25000	14	62000-63000	7
25000-26000	8	63000-64000	14
26000-27000	18	64000-65000	2
27000-28000	7	65000-66000	8
28000-29000	18	66000-67000	18
29000-30000	25	67000-68000	2
30000-31000	5	68000-69000	4
31000-32000	12	69000-70000	5
32000-33000	12	70000-80000	55
33000-34000	13	80000-90000	32
34000-35000	6	90000-100000	27
35000-36000	14	100000-150000	35
36000-37000	9	150000-200000	2
37000-38000	21		

Appendix- VI

Frequency distribution of 1025 trees based on the value of the regression estimate

Class Interval	Frequency	Class Interval	Frequency
0-9	0	27.0-27.5	17
9-10	5	27.5-28.0	13
10-10.5	4	28.0-28.5	13
10.5-11.0	10	28.5-29.0	18
11.0-11.5	10	29.0-29.5	20
11.5-12.0	8	29.5-30.0	11
12.0-12.5	22	30.0-30.5	13
12.5-13.0	10	30.5-31.0	6
13.0-13.5	62	31.0-31.5	6
13.5-14.0	31	31.5-32.0	18
14.0-14.5	18	32.0-32.5	4
14.5-15.0	32	32.5-33.0	9
15.0-15.5	15	33.0-33.5	9
15.5-16.0	18	33.5-34.0	10
16.0-16.5	17	34.0-34.5	11
16.5-17.0	17	34.5-35.0	6
17.0-17.5	29	35.0-35.5	10
17.5-18.0	24	35.5-36.0	5
18.0-18.5	15	36.0-36.5	14
18.5-19.0	20	36.5-37.0	8
19.0-19.5	14	37.0-37.5	6
19.5-20.0	17	37.5-38.0	4
20.0-20.5	16	38.0-38.5	7
20.5-21.0	24	38.5-39.0	4
21.0-21.5	19	39.0-39.5	6
21.5-22.0	15	39.5-40.0	4
22.0-22.5	19	40-45	42
22.5-23.0	22	45-50	31
23.0-23.5	17	50-55	20
23.5-24.0	22	55-60	9
24.0-24.5	13	60-65	10
24.5-25.0	23	65-70	1
25.0-25.5	19	70-80	3
25.5-26.0	10	80-90	2
26.0-26.5	21	90-100	0
26.5-27.0	15	100-110	2

OPTIMUM STRATIFICATION FOR YIELD ESTIMATION IN COCOA (*Theobroma cacao* L.)

By

SUNILKUMAR. G.

ABSTRACT OF A THESIS

Submitted in partial fulfilment of the requirement for the degree of

Master of Science in Agricultural Statistics

Faculty of Agriculture

Kerala Agricultural University

Department of Agricultural Statistics

COLLEGE OF HORTICULTURE

VELLANIKKARA, THRISSUR - 680 656

KERALA, INDIA

2000

ABSTRACT

Investigation on "Optimum stratification for yield estimation in cocoa (*Theobroma cacao*. L)" was carried out in the department of Agricultural Statistics, College of Horticulture, Vellanikkara during 1997-99 using data on 1025 'Forestro' variety cocoa trees from the Cadbury-KAU Co-operative Cocoa Research Project (CCRP), College of Horticulture, Kerala Agricultural University.

Four different stratification rules, viz, cumulative root 'f' rule, cumulative cube root 'f' rule, Ekman's rule and an iterative procedure for function minimisation, were used to arrive at boundary points of strata. For each of these rules stratification was carried out on average yield of yester years, canopy spread, value of HxG^2 where 'H' is height and 'G' is the girth of the trees , the first principal component derived from these variates and height and girth of the trees and regression estimate of the study character with the predictor variables mentioned above. Sampling variance of the estimator of the population mean under Neyman allocation for two to five strata situations was estimated in each case, assuming a uniform sample size of 200. Different stratification rules and stratification variables were compared using these estimates.

No single rule was found to be appropriate for all the stratification variables and for different number of strata. But in most of the cases the cumulative root 'f' rule was found to be good for smaller number of strata followed by the cumulative cube root 'f' rule. For large number of strata, the iterative procedure performed consistently well compared to all the other methods. In the case of stratification based on the canopy spread, the Ekman's method was found to be good for different numbers of strata.

Regarding the stratification variables, the average yield of yester years was found to be best followed by the regression estimate in the sense of resulting in smaller sampling variance of the estimator of the population mean. Stratification based on value of HxG^2 and the first principal component were found to be inferior.

Average yield of yester years or regression estimate of yield could be used as covariate to perform the analysis of covariance for experiments in cocoa and also blocking of trees could be done based on these for the conduct of planned experiments on cocoa.