COENOSES 10(2-3):95-111, 1995.

ESTIMATION OF MULTIPLE CHARACTERISTICS BY RANKED SET SAMPLING METHODS[1]

Ruthanna C. Norris, G. P. Patil, and Arun K. Sinha

Center for Statistical Ecology and Environmental Statistics, Pennsylvania State University, University Park, PA 16802, USA

Keywords: Concomitant variable, Environmental sampling, Ranked set sampling, Simple random sampling, Stratified sampling, Unequal allocation.

Abstract. In this paper we examine some ranked set sampling (RSS) methods to estimate the means of multiple characteristics in a single investigation. These methods include RSS as developed by McIntyre with equal as well as unequal allocation, and a modified version of Takahasi's RSS method. For determining the unequal allocation we use the Neyman's criterion with concomitant ranking as suggested by Sarndal, Swensson and Wretman (1992) for obtaining a near-optimal allocation for stratified sampling. In the modified Takahasi's method ranking information is utilized at the estimation stage, and as such on ignoring the information one obtains a random sample. This method is also based on the same assumption as the McIntyre's RSS method that the ranking of a unit in a small group with respect to the characteristic of interest is easier than its quantification. These procedures are illustrated using a real data set consisting of height, diameter at breast height, and age of trees. All the three methods provide better estimators of the population mean than the simple random sampling (SRS) method with the same sample size. But the McIntyre's RSS method with the unequal allocation performs better than other RSS methods under consideration.

Introduction

While dealing with multiple characteristics in a sample survey one usually encounters two scenarios. Under the first scenario the aim is to estimate a parameter, say the population mean of a characteristic of interest, using the information about the rest of the characteristics. This procedure, in fact, dichotomizes the available characteristics into two groups viz. the main characteristic and the auxiliary characteristics. The latter are used to bring refinement to the estimation of the parameters of the former. Rao (1992) provides an overview of this area. Under the second scenario one may be interested in estimating a parameter, say the population means of all available characteristics. This situation may arise in environmental and ecological investigations more often. For this kind of sample surveys each selected unit is quantified for all characteristics of interest. The main idea behind this scheme is to accomplish observational economy. While providing some insight into this area of sampling Rao (1992) mentions briefly some classical sampling methods used for this purpose. In this paper, while considering the second scenario we investigate the performances of some ranked set sampling (RSS) methods. This method of sampling is appropriate when identification and acquisition of sampling units are inexpensive compared with their quantifications, and the randomly selected units can be ranked by visual perception or by some other crude methods with respect to the main characteristic of interest without using their exact measurements. By exploiting the ranking information this method provides a more structured sample than the simple random sampling (SRS) method does with the same sample size. Though it was developed by McIntyre (1952) for estimating pasture yields, it was used first by Halls and Dell (1966) in a field survey to estimate the weight of browse and herbage in a pine-hardwood forest of east Texas. While Patil, Sinha, and Taillie (1994a) provide a detailed and up-to-date overview of this method, Patil, Sinha, and Taillie (1992a) present an annotated bibliography of the literature on this area of sampling (see also Johnson, Patil and Sinha 1993). For estimating the population means of multiple characteristics Patil, Sinha, and Taillie (1994b) discuss this method, and using two characteristics they show that the population means can be estimated more efficiently by RSS with the equal allocation than SRS with the same sample size. But the mean of a characteristic is estimated more efficiently using a ranked set sample (RSS) obtained with perfect ranking than that with concomitant ranking. Apart from this, they also discuss a ranking model based upon the concept of size biased permutations. This model permits a wide spectrum of ranking accuracy ranging from random ranking to near perfect ranking. In fact, this allows the ranking of units to depend upon several or all characteristics collective-

1 Prepared with partial support from the Statistical Analysis and Computing Branch, Environmental Statistics and Information Division, Office of Policy, Planning, and Evaluation, United States Environmental Protection Agency, Washington, DC under a Cooperative Agreement Number CR-821531. The contents have not been subjected to Agency review and therefore do not necessarily reflect the views of the Agency and no official endorsement should be inferred.

ly. While extending the work further we, in this paper, attempt to examine the performances of RSS with the equal and unequal allocation as proposed by McIntyre (1952), and a modified version of the RSS method developed by Takahasi (1970). The unequal allocation is based on the Neyman's criterion which suggests that the sample size for each rank order should be proportional to the standard deviation of that rank order. But this method requires the availability of the population standard deviation of each rank order for all characteristcs of interest. To deal with this impasse we take help of a concomitant variable. Note that it is used for providing ranks when units cannot be ranked with respect to the main characteristic of interest. As this variable is supposed to be known we use its standard deviations for various order statistics for determining the unequal allocation. In absence of these population standard deviations one may take help of some previous surveys of similar nature or one may conduct a sample survey of a small size using RSS with the equal allocation. In fact, this predicament also arises in stratified sampling while determining the optimum allocation based on the Neyman's criterion. For this situation Sarndal, Swensson, and Wretman (1992) suggest to replace the population standard deviation of the main variable by that of an auxiliary variable. They call this as ''x-optimal allocation'' because they denote the main variable by Y and the auxiliary variable by X. Depending upon the correlation between the two variables one can get a near-optimal allocation. The main benefit of the Takahasi's method is that one gets a random sample on ignoring the ranking information, and thus one can address issues that might arise retrospectively with this RSS. Also, the accuracy of existing ranks can be improved as the information base grows. For these reasons the RSS method appears to be more appropriate for environmental and ecological sampling and monitoring situations. In this paper we also introduce a modification to the RSS method developed by Takahasi (1970) for estimating multiple characteristics. The modification to the Takahasi's method ensures that there will be at least two quantifications for each rank which, in turn, allow estimates of the variance of the means to be calculated. As a result of this procedure each characteristic is estimated with its own ranking capability, and as such it may provide a better estimate than the one based on a concomitant ranking using the McIntyre's equal allocation technique. This method uses the ranking information at the estimation stage whereas the McIntyre's method applies this information for selecting units. For a large sample size the Takahasi's method shows the same performance as that of the McIntyre's RSS method with the equal allocation under perfect ranking scenario. For illustration we use a real data set referred to by Platt, Evans, and Rathbun (1988).

Methods of ranked set sampling

For estimating the means of multiple characteristics we examine RSS as proposed by McIntyre with the equal and unequal allocation, and a modified Takahasi (1970)'s RSS method. All these methods are described separately in the following sections.

RSS with equal allocation

Let us first discuss RSS as developed by McIntyre (1952). According to this method m^2 units are randomly selected from an infinite population, and the selected units are arranged in m sets, each consisting of m units. After ranking the units of each set separately with respect to the variable of interest, the unit with the lowest rank i.e. with rank one is quantified from the first set, the unit with the second lowest rank i.e. with rank two is measured from the second set, and the process is continued until the unit possessing rank m is quantified from the mth set. This process results in m quantifications out of the m^2 units selected. In order to get a larger sample size n the whole process is repeated r times. This provides a ranked set sample of size $n = mr$. We call m as the set size and r as the number of cycles. As one gets the same number of quantifications r for each rank order this procedure is called RSS with the equal allocation (RSSWEA).

Let us denote the ith order statistic from the ith sample in the jth cycle by $Y_{(i:m)j}$; $i=1,\ldots, m$; $j=1,\ldots, r$ assuming perfect ranking. The RSS estimator of the population mean μ_Y is defined as

$$\hat{\mu}_{Y:RSS} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{r} Y_{(i:m)j}}{mr} \quad (1)$$

$$var(\hat{\mu}_{Y:RSS}) = \frac{1}{m^2}\sum_{i=1}^{m}\frac{\sigma^2_{Y(i:m)}}{r}. \quad (2)$$

where $\sigma^2_{Y(i:m)}$ denotes the variance of $Y_{(i:m)}$.

In order to compare the effectiveness of RSS with SRS we compute the relative precision $RP(Y)$ of the RSS estimator as compared with the corresponding SRS estimator $\bar{Y}$. The expression for the RP is given below:

$$RP(Y) = \frac{\mathrm{var}(\bar{Y})}{\mathrm{var}(\hat{\mu}_{Y:RSS})} \quad (3)$$

$$= \frac{\sigma^2_Y}{\overline{\sigma^2}_{Y(m)}} \quad (4)$$

$$\text{where } \overline{\sigma^2}_{Y(m)} = \frac{1}{m}\sum_{i=1}^{m}\sigma^2_{Y(i:m)}. \quad (5)$$

An estimator of the RP is given below:

$$\hat{RP}(Y) = \frac{\hat{\sigma}^2_{(Y:RSS)}}{\overline{\hat{\sigma}^2}_{Y(m)}}$$

where

$$\hat{\sigma}^2_{Y(m)} = \frac{1}{m}\sum_{i=1}^{m}\hat{\sigma}^2_{Y(i:m)}$$

The unbiased RSS estimator of σ^2_Y is given below:

$$\hat{\sigma}^2_{(Y:RSS)} = \frac{mr-m+1}{m^2 r(r-1)} \sum_{i=1}^{m} \sum_{j=1}^{r} (Y_{(i:m)j} - \bar{Y}_{(i:m)})^2 + \frac{\sum_{i=1}^{m} (\bar{Y}_{(i:m)} - \hat{\mu}_{Y:RSS})^2}{m} \quad (6)$$

Inserting the expressions for $\bar{Y}_{(i:m)}$ and $\hat{\mu}_{Y:RSS}$ the estimator of σ_Y^2 is obtained as:

$$\hat{\sigma}^2_{(Y:RSS)} = \frac{mr-m+1}{m^2 r(r-1)} \sum_{i=1}^{m} \sum_{j=1}^{r} Y^2_{(i:m)j} - \frac{\sum_{i=1}^{m} T_i^2}{m^2 r^2 (r-1)} - \frac{T^2}{m^2 r^2}, \quad (7)$$

where

$$T_i = \sum_{j=1}^{r} Y_{(i:m)j} \text{ and } T = \sum_{i=1}^{m} T_i .$$

For these results see Patil, Sinha, and Taillie (1992b and 1993).

When it is not possible to rank the randomly drawn sampling units with respect to the magnitude of the variable of interest we take help of some other readily available variable for this purpose. The other variable needs to be correlated with the main variable, and is known as a concomitant variable. For example, to estimate the average volume of trees, the randomly selected trees may be ranked by their height or diameter. We use this method of ranking for drawing ranked set samples for estimating the means of multiple characteristics. Let us denote them by X, Y, and Z. Now assuming that units can be ranked relatively perfectly with respect to Y we draw a ranked set sample of size *mr* of this characteristic. These selected units are also quantified for X and Z. And the quantifications are given the same ranks as those of Y. For example, a unit whose quantification for Y gets the first rank then its quantifications for X and Z will also have the first rank. In fact, in the RSS literature this kind of ranking is called concomitant ranking. Though the RSS estimators of the means of all these characteristics are more efficient than the corresponding SRS estimators, the efficacy of an RSS estimator depends on the correlation between the variable of interest and the concomitant variable apart from the level of skewness of the concerned variable. See Patil, Sinha and Taillie (1994a). In this situation the *i*th order statistic of X from the *i*th sample in the *j*th cycle is represented by $X_{[i:m]j}$ instead of $X_{(i:m)j}$.

Using (1) we estimate the means of X and Z. But we denote the relative precisions of the RSS estimators of the population means of X and Z compared with the corresponding RSS estimators by the RP(X:Y) and the RP(Z:Y) respectively. The expression for the RP(X:Y) is given below:

$$RP\,(X:Y) = \frac{\text{var}(\bar{X})}{\text{var}(\hat{\mu}_{[X:RSS]})} \quad (8)$$

$$= \frac{\sigma_X^2}{\overline{\sigma^2}_{X[m]}} \quad (9)$$

where $\overline{\sigma^2}_{X[m]} = \frac{1}{m}\sum_{i=1}^{m} \sigma^2_{X[i:m]}$. (10)

An estimator of the RP is given below:

$$\hat{RP}(X:Y) = \frac{\hat{\sigma}^2_{[X:RSS]}}{\overline{\hat{\sigma}^2}_{X[m]}} \quad (11)$$

where $\hat{\sigma}^2_{[X:RSS]}$ is obtained using (6). Here,

$$\overline{\hat{\sigma}^2}_{X[m]} = \frac{1}{m} \sum_{i=1}^{m} \hat{\sigma}^2_{X[i:m]}$$

Likewise, we define the RP(Z:Y) and its estimator.

RSS with unequal allocation

In order to draw a ranked set sample with unequal allocation we need to know the population standard deviation of each rank order i.e. $\sigma_{(i:m)}$. Let us suppose that r_i denotes the number of sets to be quantified for the units having rank *i*. According to the Neyman's criterion

$$r_i \ \alpha \ \sigma_{(i:m)}, \quad i = 1,...,m \quad (12)$$

$$\text{or} \quad r_i = \frac{n\sigma_{(i:m)}}{\sum_{i=1}^{m} \sigma_{(i:m)}} \quad (13)$$

where $r_1 + r_2 + ... + r_m = n$. Evidently, r_i cannot be determined if $\sigma_{(i:m)}$ is not known. In this situation it is computed by using an estimator of $\sigma_{(i:m)}$. While dealing with multiple characteristics we obtain r_i using the standard deviations of order statistics of the characteristic which is used as a concomitant variable because it is supposed to be known and correlated with other characteristics. For this purpose we use the estimates of these standard deviations obtained on the basis of a small ranked set sample based on the equal allocation. In fact, the same values of r_i are used for all the characteristics under consideration. This method is also used for obtaining a near-optimal allocation for stratified sampling. See Sarndal, Swensson, and Wretman (1992, p.107).

Further, if T_i denotes the total quantifications for the units having the *i*th rank then the RSS estimator $\hat{\mu}_{(m)Yu}$ of the population mean μ_Y is given by

$$\hat{\mu}_{(m)Yu} = \frac{1}{m} \sum_{i=1}^{m} \frac{T_i}{r_i} \quad (14)$$

$$= \frac{1}{m} \sum_{i=1}^{m} \bar{Y}_{(i:m)}$$

$$\text{where} \quad \bar{Y}_{(i:m)} = \frac{T_i}{r_i}. \quad (15)$$

$\hat{\mu}_{(m)Yu}$ is an unbiased estimator of the population mean. Its variance is given below:

$$\text{var}(\hat{\mu}_{(m)Yu}) = \frac{1}{m^2} \sum_{i=1}^{m} \frac{\sigma^2_{Y(i:m)}}{r_i}. \quad (16)$$

Further, for determining the effectiveness of the RSS procedure compared with the SRS estimator the relative precision $RP(Y)$ is defined as

$$RP(Y) = \frac{\text{var}(\bar{Y})}{\text{var}(\hat{\mu}_{(m)Yu})} \tag{17}$$

$$= \left(\frac{\sigma_Y^2/n}{\sum_{i=1}^{m} \frac{\sigma_{Y(i:m)}^2}{r_i} / m^2} \right) \tag{18}$$

$$= \left(\frac{\sigma_Y}{\bar{\sigma}_Y} \right)^2 \quad \text{where} \quad \bar{\sigma}_Y = \frac{1}{m} \sum_{i=1}^{m} \sigma_{Y(i:m)}. \tag{19}$$

We may compute the $RP(Y)$ if $\sigma_{Y(i:m)}$ and σ_Y are known.

If σ_Y and $\bar{\sigma}_Y$ are not known then we use their estimators to obtain the $\hat{RP}(Y)$. For this purpose we use the estimators of σ_Y^2 and $\sigma_{Y(i:m)}^2$ based on a ranked set sample obtained by the unequal allocation. The aforementioned formulas are also used for other characteristics. For this purpose the unbiased RSS estimator of the population variance σ_Y^2 is given below:

$$\hat{\sigma}^2_{RSS(YU)} = \sum_{i=1}^{m} \left(\frac{m(r_i - 1) + 1}{m^2 r_i (r_i - 1)} \right) \sum_{j=1}^{r_i} \left(Y_{(i:m)j} - \bar{Y}_{(i:m)} \right)^2 + \frac{1}{m} \sum_{i=1}^{m} \left(\bar{Y}_{(i:m)} - \hat{\mu}_{(m)Yu} \right)^2 \tag{20}$$

$$= \sum_{i=1}^{m} \left(\frac{m(r_i - 1) + 1}{m^2 r_i (r_i - 1)} \right) \left(\sum_{j=1}^{r_i} Y^2_{(i:m)j} - \frac{T_i^2}{r_i} \right) + \frac{1}{m} \sum_{i=1}^{m} \left(\frac{T_i}{r_i} \right)^2 - \left(\frac{1}{m} \sum_{i=1}^{m} \frac{T_i}{r_i} \right)^2 \tag{21}$$

Takahasi's RSS method

According to this method a unit is randomly selected from each set, and after quantifying it a rank is given to the quantification. For estimating a population mean it is necessary that every rank has at least one quantification. With this end in view one ranked set sample of size m is obtained using the McIntyre's RSS method with the equal allocation first. In order to carry out this method for estimating the population means of multiple characteristics a selected unit is quantified separately for each characteristic, and each quantification is recorded with its rank. For estimating a population mean we use the following estimator:

$$\hat{\mu}_{T:RSS} = \frac{1}{m} \sum_{i=1}^{m} \frac{T_i}{n_i + 1}$$

where T_i denotes the sum of the quantifications with rank i and n_i represents the number of quantifications with rank i obtained by the Takahasi's method. In fact, (n_i+1) is the same as r_i in (14). Moreover,

$$\sum_{i=1}^{m} r_i = \sum_{i=1}^{m} n_i + m$$

$$\text{or,} \sum_{i=1}^{m} n_i = n - m$$

The variance of the estimator is given below:

$$\text{var}(\hat{\mu}_{T:RSS}) = \frac{1}{(n-m+1)m} \left[1 - (1 - \frac{1}{m})^{n-m+1} \right] \sum_{i=1}^{m} \sigma^2_{(i:m)} . \tag{22}$$

In order to obtain an estimate of (22) we need to estimate $\sum_{i=1}^{m} \sigma^2_{(i:m)}$, where

$$\hat{\sigma}^2_{(i:m)} = \frac{1}{n_i} \sum_{j=1}^{n_{i+1}} (Y_{(i:m)j} - \hat{\mu}_{(i:m)})^2$$

Here, $\hat{\mu}_{(i:m)}$ denotes an unbiased estimator of the population mean of the ith order statistic.

Obviously, we cannot estimate $\sigma^2_{(i:m)}$ if $n_i = 0$. As such we must have a ranked set sample using the McIntyre's method with two cycles, if we wish to obtain an estimate of $\text{var}(\hat{\mu}_{T:RSS})$. In this situation we define

$$\hat{\mu}_{T:RSS} = \frac{1}{m} \sum_{i=1}^{m} \frac{T_i}{n_i + 2} \tag{23}$$

and obtain an expression of $\text{var}(\hat{\mu}_{T:RSS})$ in the following theorem. Here, $\sum_{i=1}^{m} n_i = n - 2m$.

Theorem: If $\hat{\mu}_{T:RSS} = \frac{1}{m} \sum_{i=1}^{m} \frac{T_i}{n_i + 2}$ then

$$\text{var}(\hat{\mu}_{T:RSS}) = \frac{1}{(n-2m+1)m} \left[1 - \frac{m}{(n-2m+2)} [1 - (1 - \frac{1}{m})^{n-2m+2}] \right] \sum_{i=1}^{m} \sigma^2_{(i:m)}.$$

Proof: Let n_i denote the number of quantifications with rank i following the Takahasi's method. Here, $(n_i, \ldots, n_m)$ has a multinomial distribution. Then,

$$P[(n_1, \ldots, n_m) = (k_1, \ldots, k_m)] = \frac{h!}{\prod_{i=1}^{m} k_i} \prod_{i=1}^{m} (\frac{1}{m})^{k_i}$$

$$\text{where,} \ h = \sum_{i=1}^{m} n_i = \sum_{i=1} mk_i$$

Now,

$$\text{var}(\hat{\mu}_{T:RSS}) = E(\hat{\mu}_{T:RSS} - \mu)^2$$

$$= E[E[(\hat{\mu}_{T:RSS} - \mu)^2 \mid (n_1, \ldots, n_m) = (k_1, \ldots, k_m)]]$$

$$= E\left[\sum_{i=1}^{m} \left(\frac{1}{m}\right)^2 \frac{\sigma^2_{(i:m)}}{(k_i+2)}\right]$$

$$= \frac{1}{m^2}\sum_{i=1}^{m}\left[\sum_{k_i=0}^{h}\binom{h}{k_i}\left(\frac{1}{m}\right)^{k_i}\left(1-\frac{1}{m}\right)^{h-k_i}\left(\frac{1}{k_i+2}\right)\right]\sigma^2_{(i:m)}$$

As

$$\frac{1}{k_i+2} = \frac{1}{k_i+1} - \frac{1}{(k_i+1)(k_i+2)}$$

we get

$var(\hat{\mu}_{T:RSS})$ =

$$\frac{1}{m^2}\left[\sum_{i=1}^{m}\left[\sum_{k_i=0}^{h}\binom{h}{k_i}\left(\frac{1}{m}\right)^{k_i}\left(1-\frac{1}{m}\right)^{h-k_i}\left(\frac{1}{k_i+1}-\frac{1}{(k_i+1)(k_i+2)}\right)\right]\right]\sum_{i=1}^{m}\sigma^2_{(i:m)}$$

$$= \left(\frac{1}{h+1}\right)\left(\frac{1}{m}\right)\sum_{i=1}^{m}\left[1-\left(1-\frac{1}{m}\right)^{h+1}\right]\sigma^2_{(i:m)}$$

$$- \frac{1}{(h+1)(h+2)}\sum_{i=1}^{m}\left[1-(h+2)\frac{1}{m}\left(1-\frac{1}{m}\right)^{h+1} - \left(1-\frac{1}{m}\right)^{h+2}\right]\sum_{i=1}^{m}\sigma^2_{(i:m)}$$

$$= \frac{1}{m(h+1)}\left[1-\left(\frac{m}{h+2}\right)\left[1-\left(1-\frac{1}{m}\right)^{h+2}\right]\right]\sum_{i=1}^{m}\sigma^2_{(i:m)}$$

On putting $h=n-2m$ we obtain

$var(\hat{\mu}_{T:RSS})$ =

$$\frac{1}{(n-2m+1)m}\left[1-\frac{m}{(n-2m+2)}\left[1-\left(1-\frac{1}{m}\right)^{n-2m+2}\right]\right]\sum_{i=1}^{m}\sigma^2_{(i:m)}. \quad (24)$$

The expression for the $RP(Y)$ is given below:

$$RP(Y) = \frac{var(Y)}{var(\hat{\mu}_{T:RSS})}. \quad (25)$$

Illustration

In order to examine the performances of these RSS methods for estimating the means of multiple characteristics we use a real data set consisting of height (Y), diameter (X) at breast height, and age (Z) of 399 trees. See Platt, Evans and Rathbun (1988) for a detailed description of the data set. We obtain the following result for a specific sample with the set size 6 and the number of cycles 10.

With the equal allocation we get an estimate of the *RP* as 3.40, 2.36, and 1.76 for height, diameter and age respectively whereas the corresponding values in the case of the unequal allocation are 4.21, 3.05, and 2.60 respectively. While using the modified Takahasi's method we obtain the estimates of the RP's for height, diameter, and age as 3.17, 2.95,and 2.86 respectively. This one sample illustration suggests that the unequal allocation is overall the most efficient method of RSS. The modified Takahasi's method appears to be better than the McIntyre's method based on the equal allocation for estimating the means of the variables which use concomitant rankings. The estimates of the population means, variances of the estimators, and the relative precisions obtained from 2000 simulated trials based on the real data set follow and conform to our original intuitions concerning the efficiencies of the three RSS methods.

We have chosen height as the concomitant variable, the variable upon which ranking is based. In a retrospective study such as this one, we have the benefit of perfect ranking according to height. In practice, however, this ranking is done without using the exact measurements, and therefore is subject to judgement errors. Consequently, the performance of the RSS methods in this study may be inflated. The number of quantifications has been fixed at n=60 with possible values for m of 1,2,...,6, where m=1 corresponds to SRS. Also, note that the samples are drawn with replacement as not to diminish the variability present in the parent population. Further,in order to have an idea of the relative gain due to an RSS estimator as compared to the simple random sample (SRS) estimator the relative saving (RS) is computed. It is obtained by subtracting the reciprocal of the RP from one. With a view to examine the variability of the characteristics under consideration the coefficient of variation (CV), the ratio of standard deviation to mean, is computed. These values are expressed in percentages.

Simulations with equal allocation

We have simulated RSS with the equal allocation trials as discussed in the Section on RSS with equal allocation and calculated the estimators $\hat{\mu}_{Y:RSS}$, $\hat{\sigma}^2_{(Y:RSS)}$, $\hat{var}(\hat{\mu}_{Y:RSS})$, and$\hat{RP}(Y)$ using equations (1), (7), (2), and (11) respectively for each trial. The averages of the estimates obtained from the 2000 trials are given in Table 1. The $var(\hat{\mu}_{Y:RSS})$ was also estimated by the sample variance of the 2000 simulated values of $\hat{\mu}_{Y:RSS}$ and is denoted by $\hat{var}(\hat{\mu}_{Y:RSSample})$.

There are two important patterns which are revealed by Table 1. First, the relative precision estimates for all sizes of m are the highest for height, while the $\hat{RP}(X:Y)$'s fall second and the $\hat{RP}(Z:Y)$'s are the smallest. This is due to a fact that we had perfect ranking on the variable Y and concomitant rankings for the X and the Z. Moreover, the correlation between Y and X is higher than the correlation between Y and Z causing the$\hat{RP}(X:Y)$'s to be higher than the $\hat{RP}(Z:Y)$'s. The other factor which is responsible for this finding is the level of skewness which is higher for age than that for diameter. Also revealed by the table is the relationship between the set size m, and the variance of $\hat{\mu}_{RSS}$. More explicitly, as m becomes larger, the estimate of $\hat{\mu}_{RSS}$ becomes more precise. The histograms of the sampling distributions of $\hat{\mu}_{RSS}$, $\hat{var}(\hat{\mu}_{RSS})$, and $\hat{cv}(RSS)$ in Figures 1, 2 and 3, respectively, graphically display this result.

Simulations with unequal allocation

Simulations using the unequal allocation as discussed in the Section on RSS with unequal allocation were carried out along with calculations of $\hat{\mu}_{Y:RSS}$, $\hat{\sigma}^2_{(Y:RSS)}$, $\hat{var}(\hat{\mu}_{Y:RSS})$, and$\hat{RP}(Y)$ using equations (14), (21), (16), and (19) respec-

Table 1. Simulation results for ranked set sampling with the equal allocation

Y=Height							
m	$\hat{\mu}_{Y:RSS}$	$\widehat{\sigma^2}_{(Y:RSS)}$	$\widehat{var}(\hat{\mu}_{Y:RSS})$	$\widehat{var}(\hat{\mu}_{Y:RSSample})$	$\widehat{cv}(Y)$	$\widehat{RP}(Y)$	$\widehat{RS}(Y)$
1	51.83	958.77	15.98	16.14	59.88	1.00	0.00
2	51.58	958.91	10.76	10.61	60.08	1.51	33.87
3	51.69	959.75	8.10	7.93	59.93	2.02	50.60
4	51.75	960.46	6.53	6.29	59.87	2.53	60.53
5	51.64	961.72	5.46	5.65	60.04	3.04	67.16
6	51.62	961.41	4.65	4.39	60.03	3.57	72.00

X=Diameter							
m	$\hat{\mu}_{X:RSS}$	$\widehat{\sigma^2}_{(X:RSS)}$	$\widehat{var}(\hat{\mu}_{X:RSS})$	$\widehat{var}(\hat{\mu}_{X:RSSample})$	$\widehat{cv}(X:Y)$	$\widehat{RP}(X:Y)$	$\widehat{RS}(X:Y)$
1	20.99	310.91	5.18	5.26	84.21	1.00	0.00
2	20.85	309.68	3.77	3.82	84.45	1.39	28.07
3	20.89	309.68	3.01	3.00	84.16	1.76	43.12
4	20.92	311.38	2.57	2.46	84.28	2.10	52.27
5	20.87	310.33	2.26	2.39	84.30	2.38	58.07
6	20.87	310.58	2.01	1.90	84.31	2.69	62.89

Z=Age							
m	$\hat{\mu}_{Z:RSS}$	$\widehat{\sigma^2}_{(Z:RSS)}$	$\widehat{var}(\hat{\mu}_{Z:RSS})$	$\widehat{var}(\hat{\mu}_{Z:RSSample})$	$\widehat{cv}(Z:Y)$	$\widehat{RP}(Z:Y)$	$\widehat{RS}(Z:Y)$
1	52.94	3282.00	54.70	56.14	107.84	1.00	0.00
2	52.33	3229.38	42.90	43.95	108.00	1.27	20.96
3	52.56	3255.60	37.00	36.80	107.88	1.49	32.88
4	52.49	3248.79	33.17	33.33	107.84	1.68	40.35
5	52.55	3265.97	30.84	31.83	108.04	1.82	45.02
6	52.51	3262.55	28.44	27.02	108.06	1.91	47.70

Table 2. Simulation results for ranked set sampling with the unequal allocation

Y=Height							
m	$\hat{\mu}_{Y:RSS}$	$\widehat{\sigma^2}_{(Y:RSS)}$	$\widehat{var}(\hat{\mu}_{Y:RSS})$	$\widehat{var}(\hat{\mu}_{Y:RSSample})$	$\widehat{cv}(Y)$	$\widehat{RP}(Y)$	$\widehat{RS}(Y)$
1	51.54	960.79	16.01	15.55	60.27	1.00	0.00
2	51.74	959.14	10.74	11.52	59.95	1.52	34.38
3	51.61	956.48	8.09	7.95	59.96	2.05	51.16
4	51.71	960.81	6.51	6.84	59.98	2.59	61.38
5	51.59	960.70	5.44	5.58	60.09	3.14	68.10
6	51.67	959.94	4.66	4.25	59.97	3.65	72.60

X=Diameter							
m	$\hat{\mu}_{X:RSS}$	$\widehat{\sigma^2}_{(X:RSS)}$	$\widehat{var}(\hat{\mu}_{X:RSS})$	$\widehat{var}(\hat{\mu}_{X:RSSample})$	$\widehat{cv}(X:Y)$	$\widehat{RP}(X:Y)$	$\widehat{RS}(X:Y)$
1	20.80	309.41	5.16	4.92	84.74	1.00	0.00
2	20.92	309.80	3.57	3.89	84.24	1.49	32.62
3	20.84	308.22	2.79	2.71	84.26	1.91	47.74
4	20.88	309.39	2.31	2.39	84.22	2.34	57.24
5	20.84	309.44	1.96	2.01	84.37	2.77	63.88
6	20.88	309.81	1.75	1.60	84.23	3.09	67.68

Z=Age							
m	$\hat{\mu}_{Z:RSS}$	$\widehat{\sigma^2}_{(Z:RSS)}$	$\widehat{var}(\hat{\mu}_{Z:RSS})$	$\widehat{var}(\hat{\mu}_{Z:RSSample})$	$\widehat{cv}(Z:Y)$	$\widehat{RP}(Z:Y)$	$\widehat{RS}(Z:Y)$
1	52.41	3258.99	54.32	53.31	108.46	1.00	0.00
2	52.50	3233.97	39.54	42.26	107.88	1.40	28.56
3	52.27	3224.78	32.96	32.85	108.03	1.68	40.43
4	52.49	3241.24	28.69	28.94	107.99	1.94	48.53
5	52.43	3252.36	25.84	26.83	108.22	2.17	53.83
6	52.50	3250.87	24.20	23.70	107.93	2.32	56.83

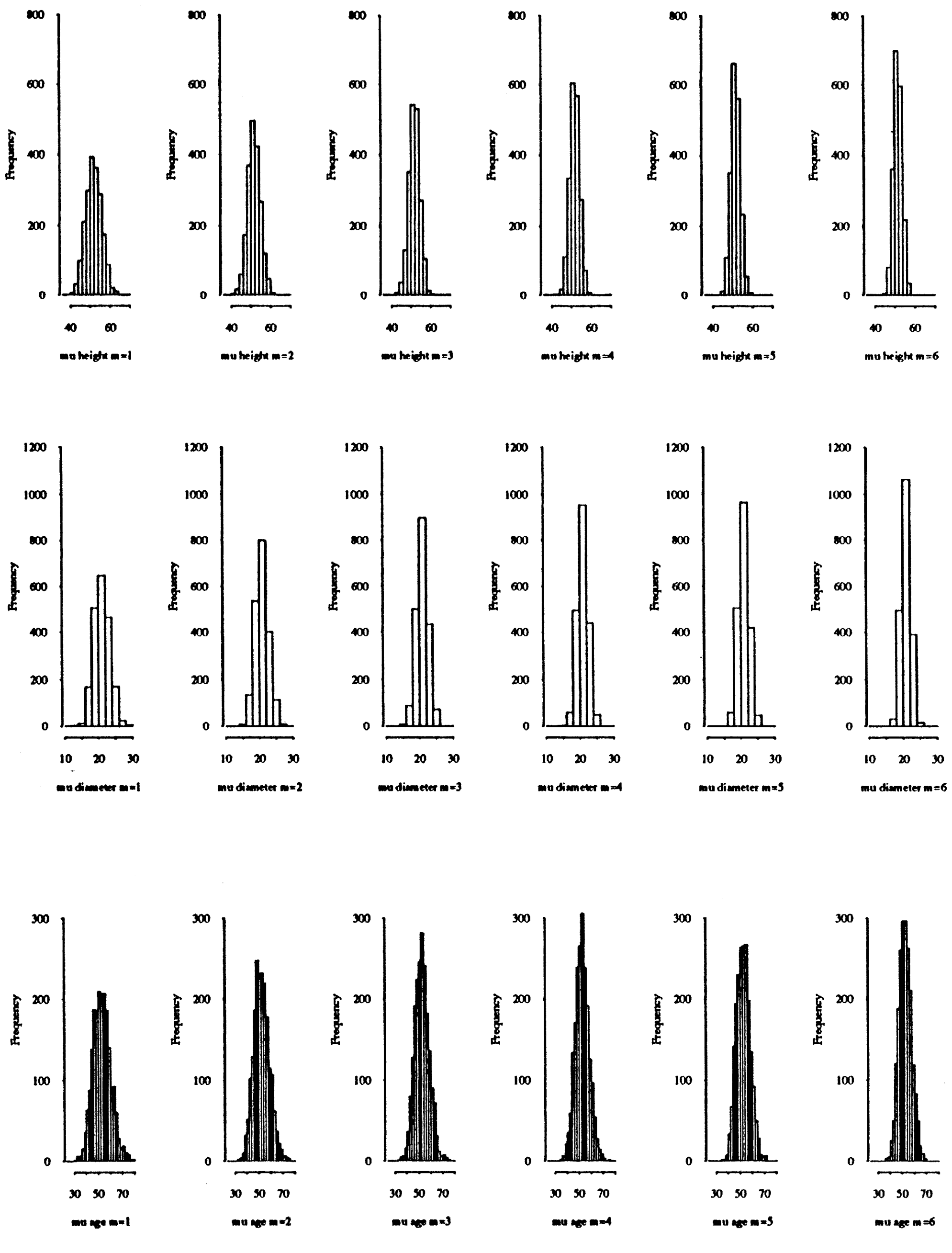

Figure 1. Sampling distribution of $\hat{\mu}_{RSS}$ for height, diameter, and age while using the equal allocation.

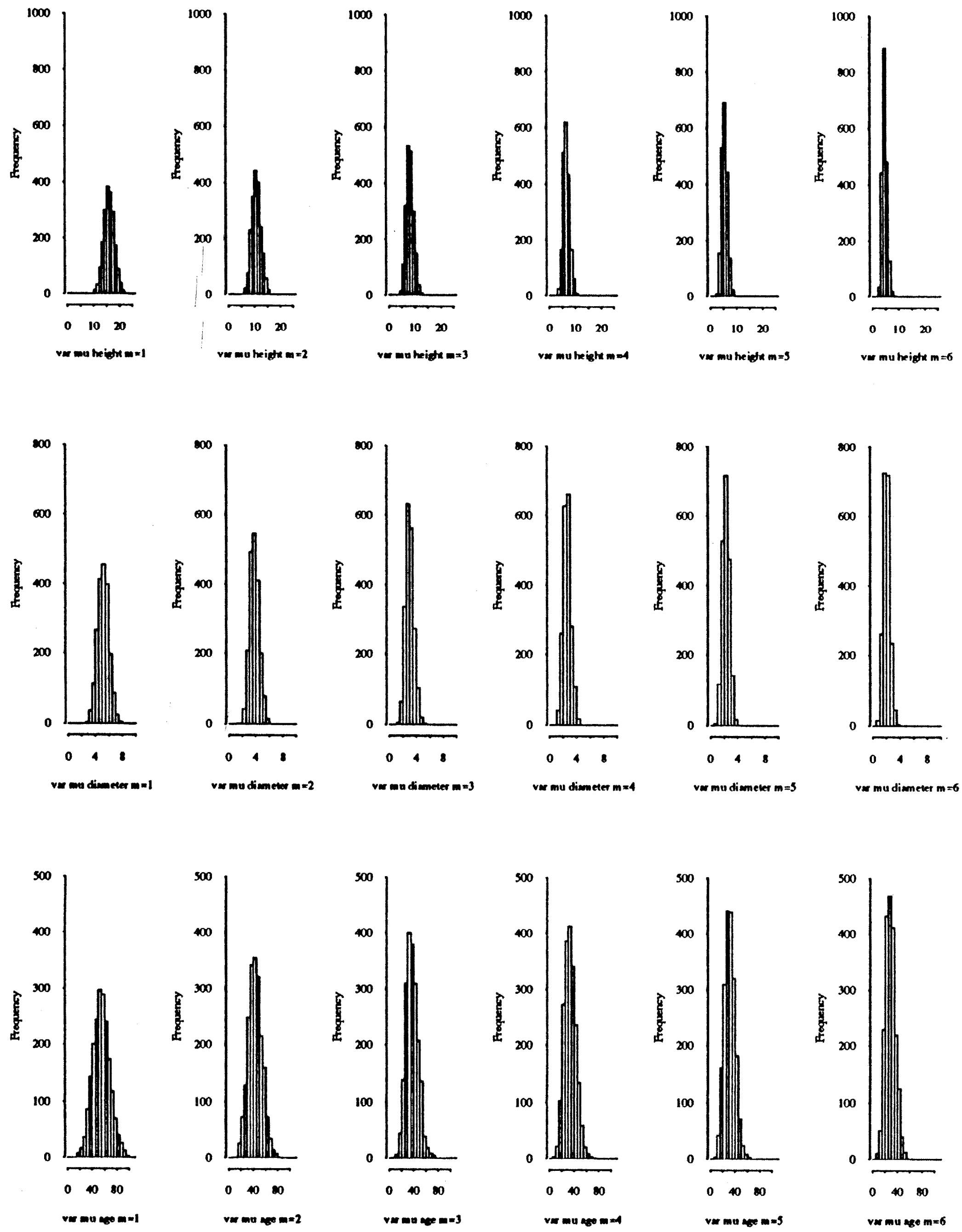

Figure 2. Sampling distribution of $\widehat{\text{var}}\,(\hat{\mu}_{RSS})$ for height, diameter, and age while using the equal allocation.

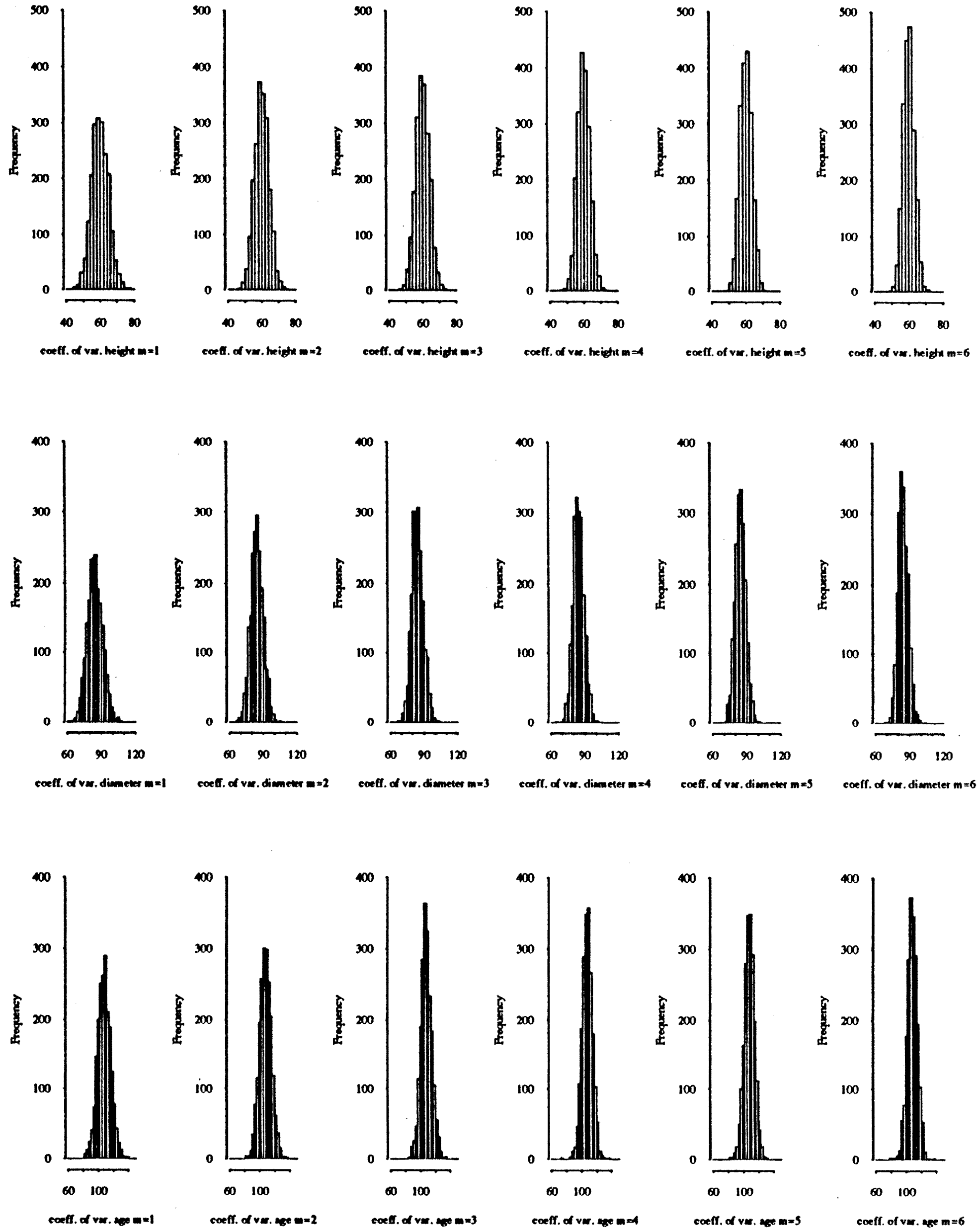

Figure 3. Sampling distribution of $\widehat{cv}(RSS)$ for height, diameter, and age while using the equal allocation.

tively for for each trial. The averages over 2000 trials are reported in Table 2. As with the equal allocation, the $\widehat{var}(\hat{\mu}_{Y:RSS})$ was also estimated by the sample variance of the 2000 simulated values of $\hat{\mu}_{Y:RSS}$.

The two patterns which emerged in the case of the equal allocation surface again while using the unequal allocation. First, the variable on which ranking is perfect has the highest relative precision. For other variables, the magnitude of the relative precision increases (decreases) as the correlation between the other variable and the ranking variable increases (decreases). As mentioned earlier the skewness of a variable is also responsible for this difference. And second, as m becomes larger, the estimate of $\hat{\mu}_{Y:RSS}$ becomes more precise. Figures 4, 5 and 6 graphically reinforce this result.

Simulations with the modified Takahasi's method

Simulations of the modified Takahasi's method as described in the Section on Takahasi's RSS method and the corresponding estimates of $\hat{\mu}_{Y:RSS}$, $\hat{\sigma}^2_{(Y:RSS)}$, $\widehat{var}(\hat{\mu}_{Y:RSS})$, and $\widehat{RP}(Y)$ using equations (23), (21), (24), and (25) respectively for each of the 2000 trials yield results similar to those of the equal and the unequal allocation. First, the relative precisions are the highest for height. And second, as m increases the $\widehat{var}(\hat{\mu}_{Y:RSS})$ decreases. The estimates are shown in Table 3.

Figures 7, 8 and 9 display the sampling distributions of $\hat{\mu}_{Y:RSS}$, $\widehat{var}(\hat{\mu}_{RSS})$, and $\widehat{cv}(RSS)$, respectively.

Conclusion

The intent of this paper is to examine the efficiencies of RSS methods which include the McIntyre's RSS methods based on the equal and unequal allocations, and a modified Takahasi's method as compared with the SRS method for estimating the population means of multiple characteristics. Considering the RP estimates for each of these methods given in Tables 1 through 3, we draw conclusions about which RSS method is most effective and when.

Under the scenario of perfect ranking i.e. for height the performance of the McIntyre's RSS method with the unequal allocation is the best followed by the RSS method with the equal allocation, and finally the modified Takahasi's method. For diameter and age also, the case of imperfect ranking, we find that the unequal allocation performs the best. In this situation the modified Takahasi's method shows better performance than the McIntyre's method with the equal allocation because the former is based on perfect ranking whereas the latter utilizes concomitant ranking. Moreover, the higher values of the RP's for diameter than age are observed because of lower variability and skewness of the former than those of the latter. Note that the estimates of the CV concentrate around 60%, 84%, and 108% for height, diameter, and age respectively.

While selecting an RSS method for estimating the population means of multiple characteristics one may prefer the McIntyre's RSS method with the unequal allocation to

Table 3. Simulation results for ranked set sampling with the modified Takahasi's method

Y=Height							
m	$\hat{\mu}_{Y:RSS}$	$\widehat{\sigma^2}_{(Y:RSS)}$	$\widehat{var}(\hat{\mu}_{Y:RSS})$	$\widehat{var}(\hat{\mu}_{Y:RSSample})$	$\widehat{cv}(Y)$	$\widehat{RP}(Y)$	$\widehat{RS}(Y)$
1	51.82	958.31	15.97	15.53	59.86	1.00	0.00
2	51.84	957.00	10.94	11.53	59.71	1.49	32.66
3	51.69	956.45	8.34	8.29	59.84	1.96	49.06
4	51.65	957.82	6.83	6.85	59.91	2.41	58.56
5	51.65	955.95	5.73	5.77	59.85	2.88	65.31
6	51.70	959.65	5.04	4.69	59.89	3.31	69.81

X=Diameter							
m	$\hat{\mu}_{X:RSS}$	$\widehat{\sigma^2}_{(X:RSS)}$	$\widehat{var}(\hat{\mu}_{X:RSS})$	$\widehat{var}(\hat{\mu}_{X:RSSample})$	$\widehat{cv}(X:Y)$	$\widehat{RP}(X:Y)$	$\widehat{RS}(X:Y)$
1	20.91	310.07	5.17	5.05	84.40	1.00	0.00
2	20.96	310.44	3.73	3.83	84.10	1.40	28.79
3	20.82	309.07	2.91	2.86	84.38	1.82	45.00
4	20.87	309.48	2.47	2.44	84.22	2.16	53.75
5	20.84	308.90	2.13	2.07	84.25	2.52	60.35
6	20.90	309.70	1.92	1.90	84.11	2.84	64.73

Z=Age							
m	$\hat{\mu}_{Z:RSS}$	$\widehat{\sigma^2}_{(Z:RSS)}$	$\widehat{var}(\hat{\mu}_{Z:RSS})$	$\widehat{var}(\hat{\mu}_{Z:RSSample})$	$\widehat{cv}(Z:Y)$	$\widehat{RP}(Z:Y)$	$\widehat{RS}(Z:Y)$
1	52.50	3243.41	54.06	52.16	108.10	1.00	0.00
2	52.63	3241.62	41.59	42.27	107.62	1.30	23.20
3	52.40	3239.68	34.56	34.64	107.94	1.58	36.75
4	52.35	3254.36	30.31	30.66	108.23	1.83	45.29
5	52.44	3254.83	27.19	27.44	108.11	2.07	51.66
6	52.54	3256.41	25.06	25.28	107.93	2.27	55.97

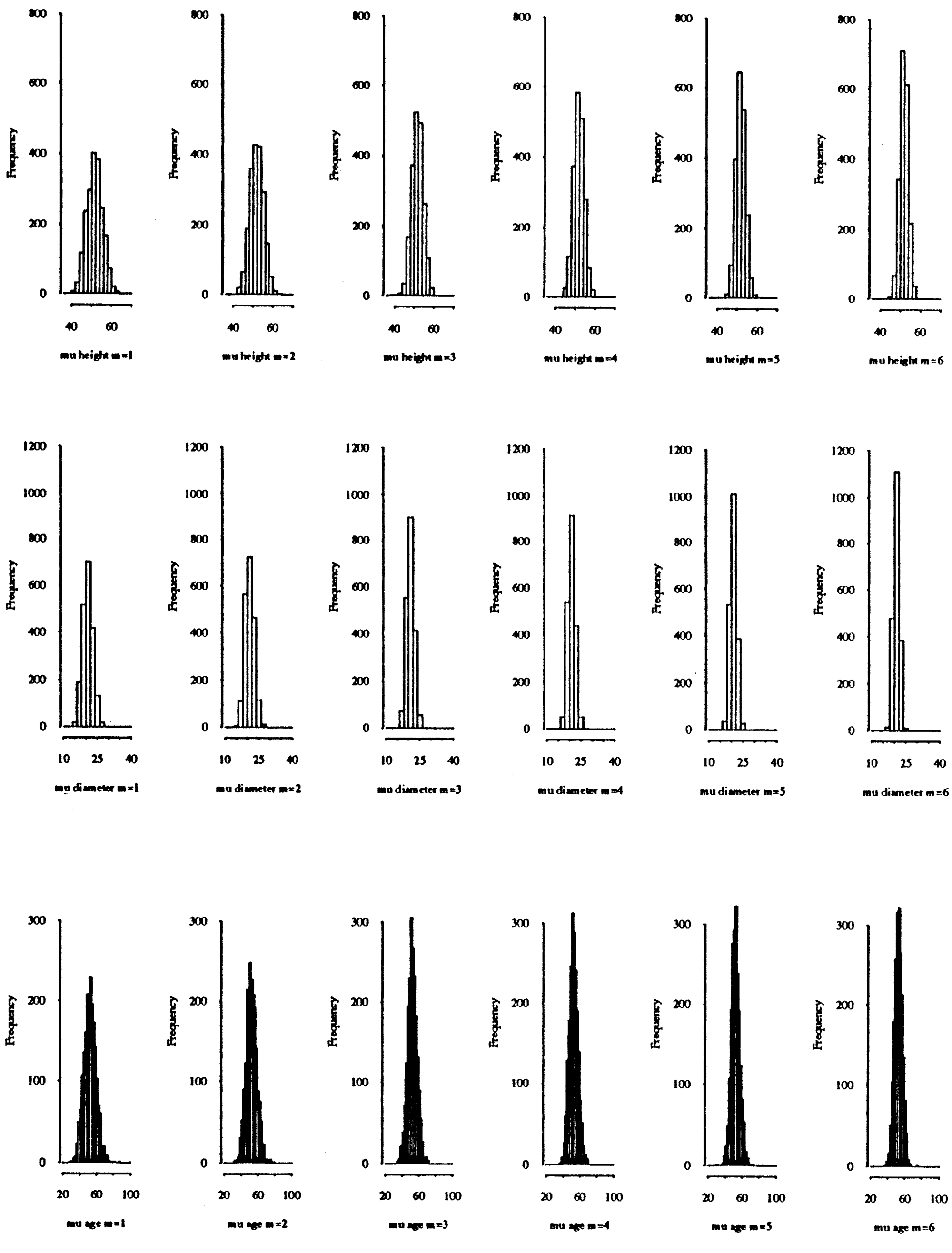

Figure 4. Sampling distribution of $\hat{\mu}_{RSS}$ for height, diameter, and age while using the unequal allocation.

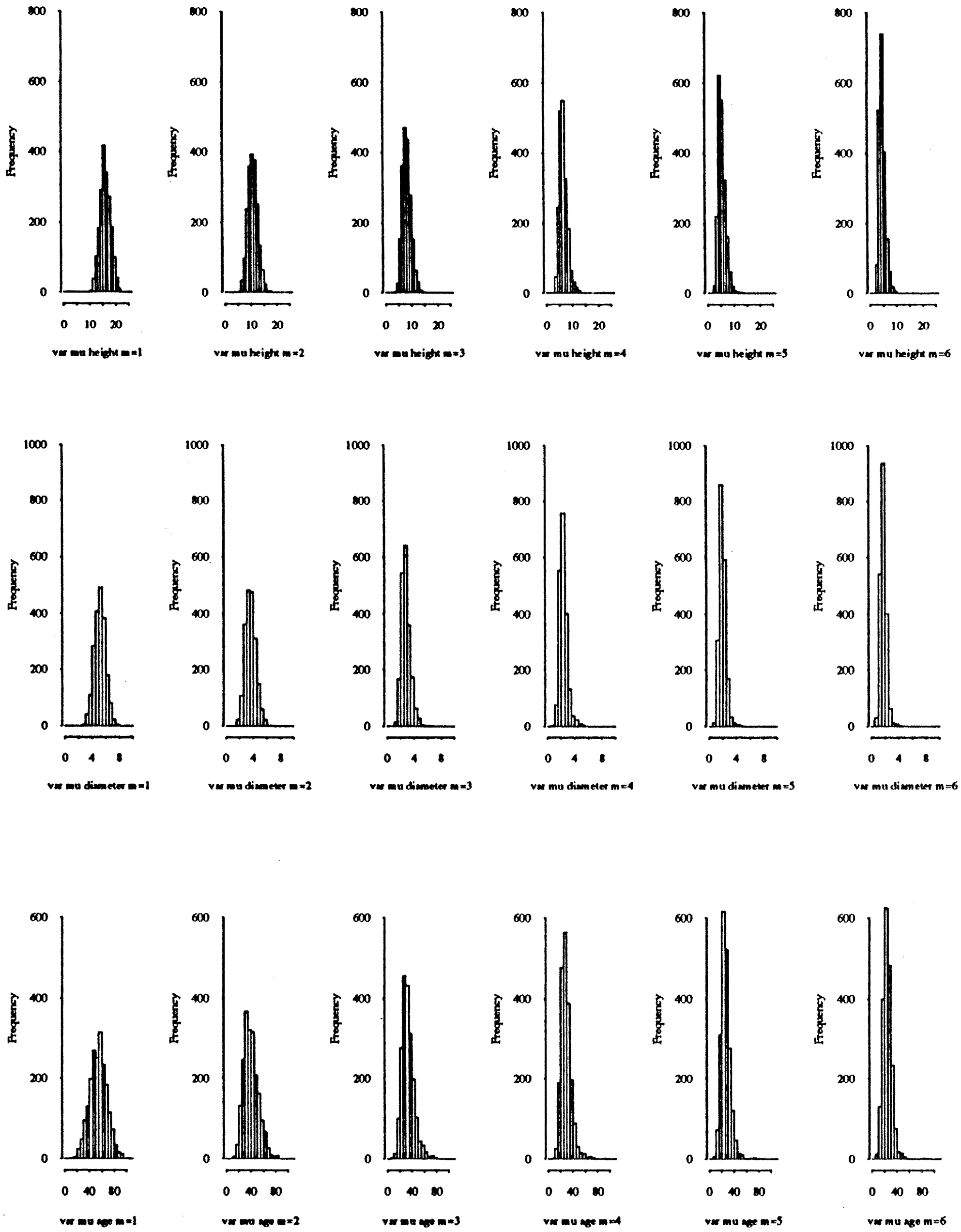

Figure 5. Sampling distribution of $\widehat{\text{var}}\,(\hat{\mu}_{RSS})$ for height, diameter, and age while using the unequal allocation.

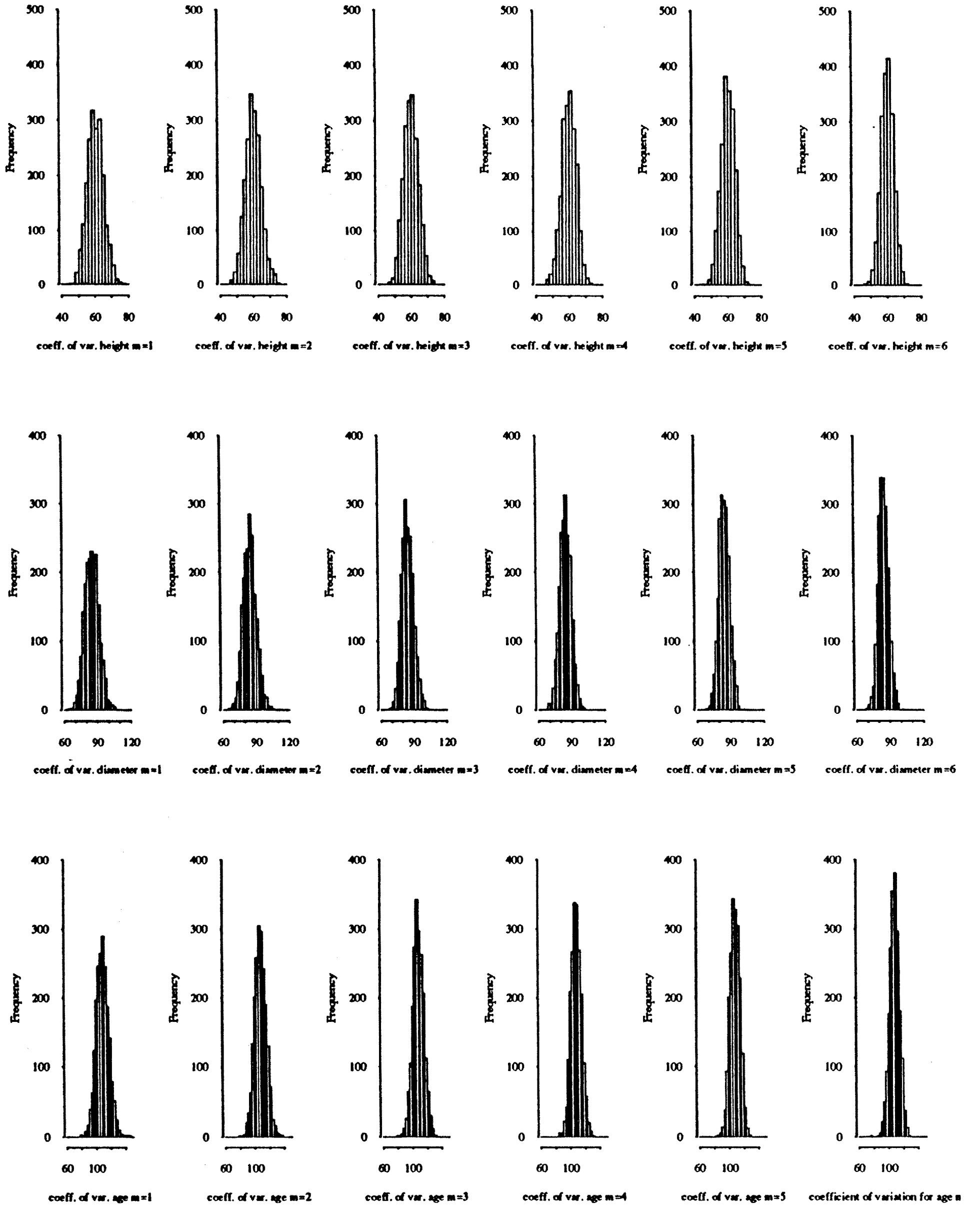

Figure 6. Sampling distribution of $\widehat{cv}(RSS)$ for height, diameter, and age while using the unequal allocation.

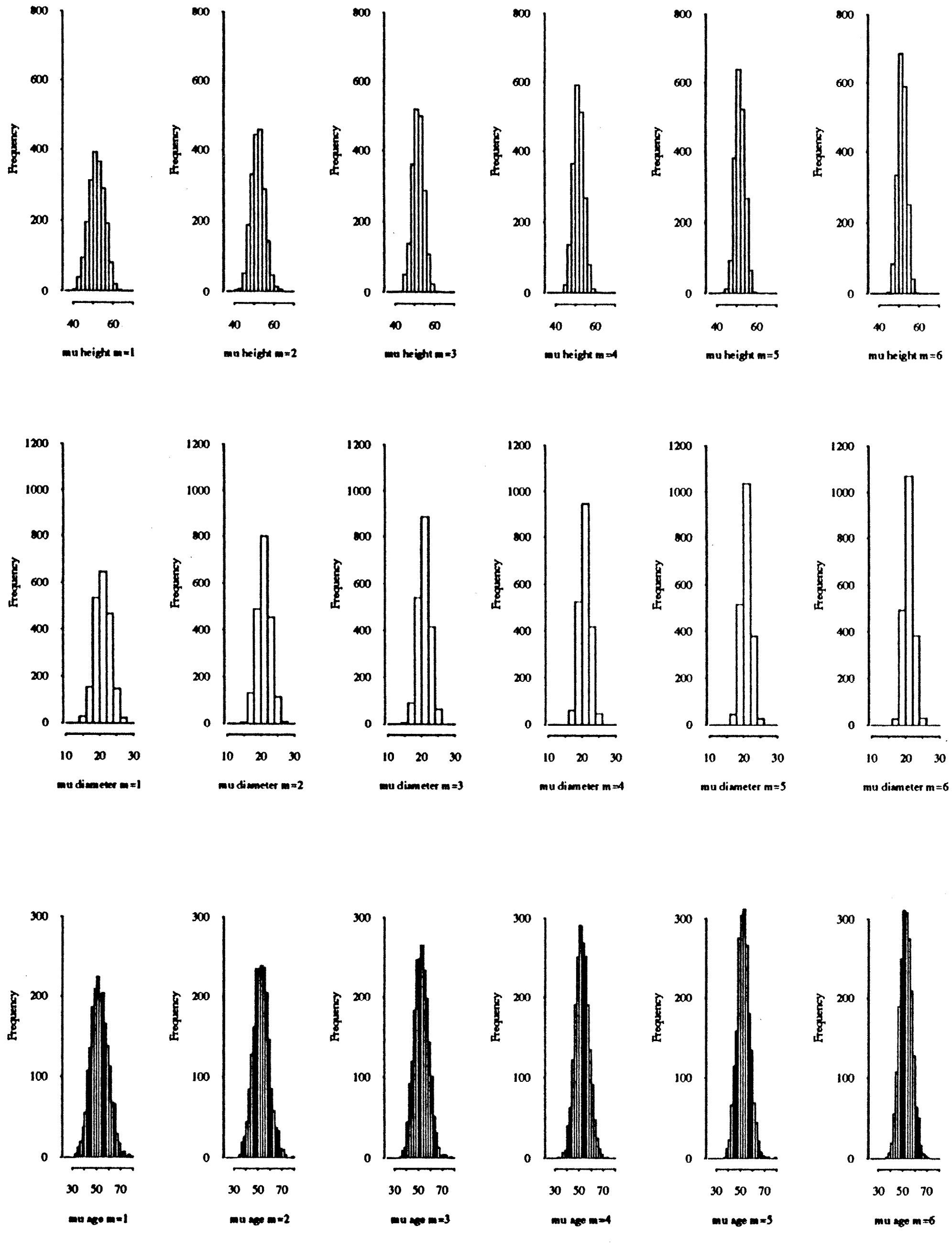

Figure 7. Sampling distribution of $\hat{\mu}_{RSS}$ for height, diameter, and age while using the modified Takahasi's method.

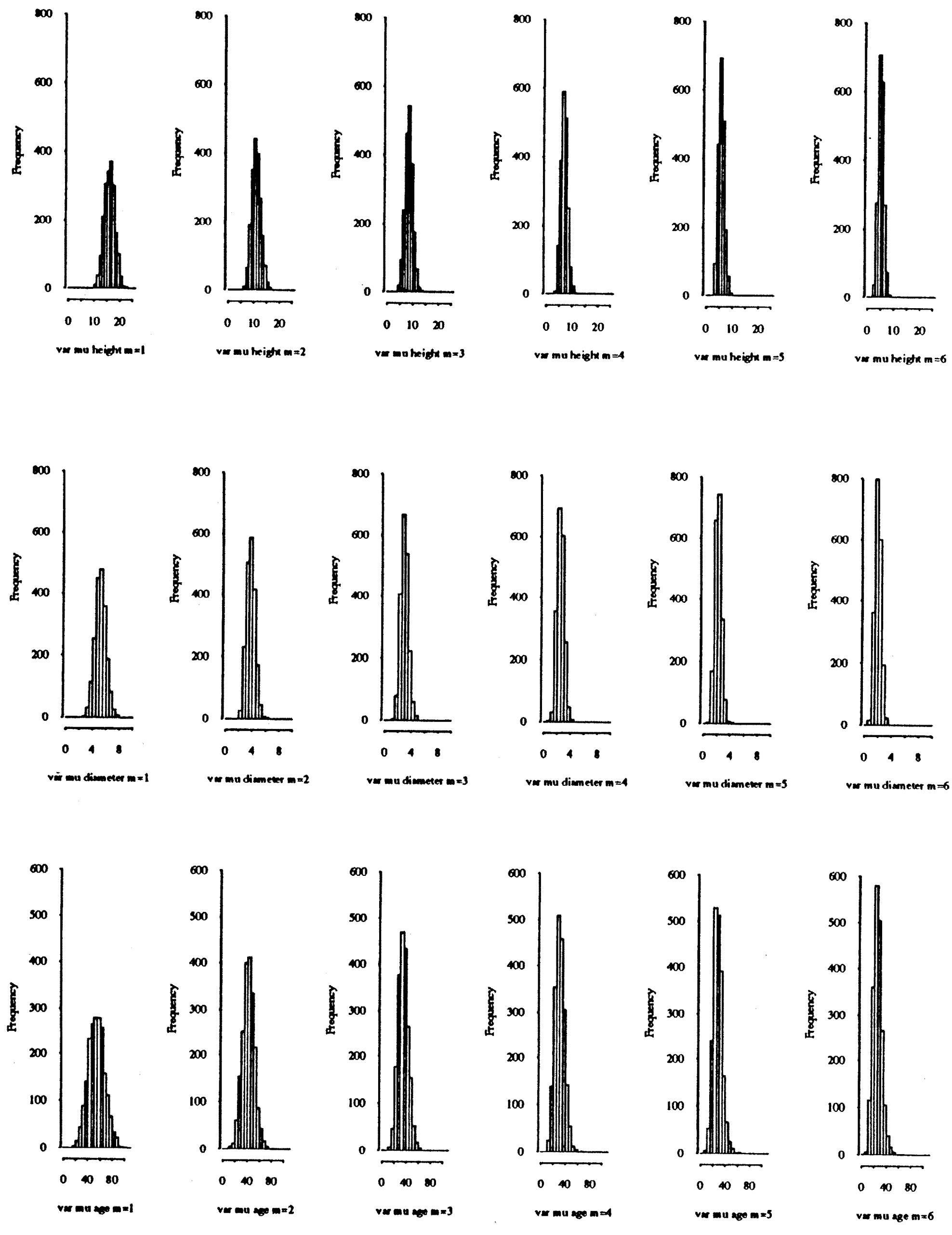

Figure 8. Sampling distribution of $\widehat{\text{var}}\ (\hat{\mu}_{RSS})$ for height, diameter, and age while using the modified Takahasi's method.

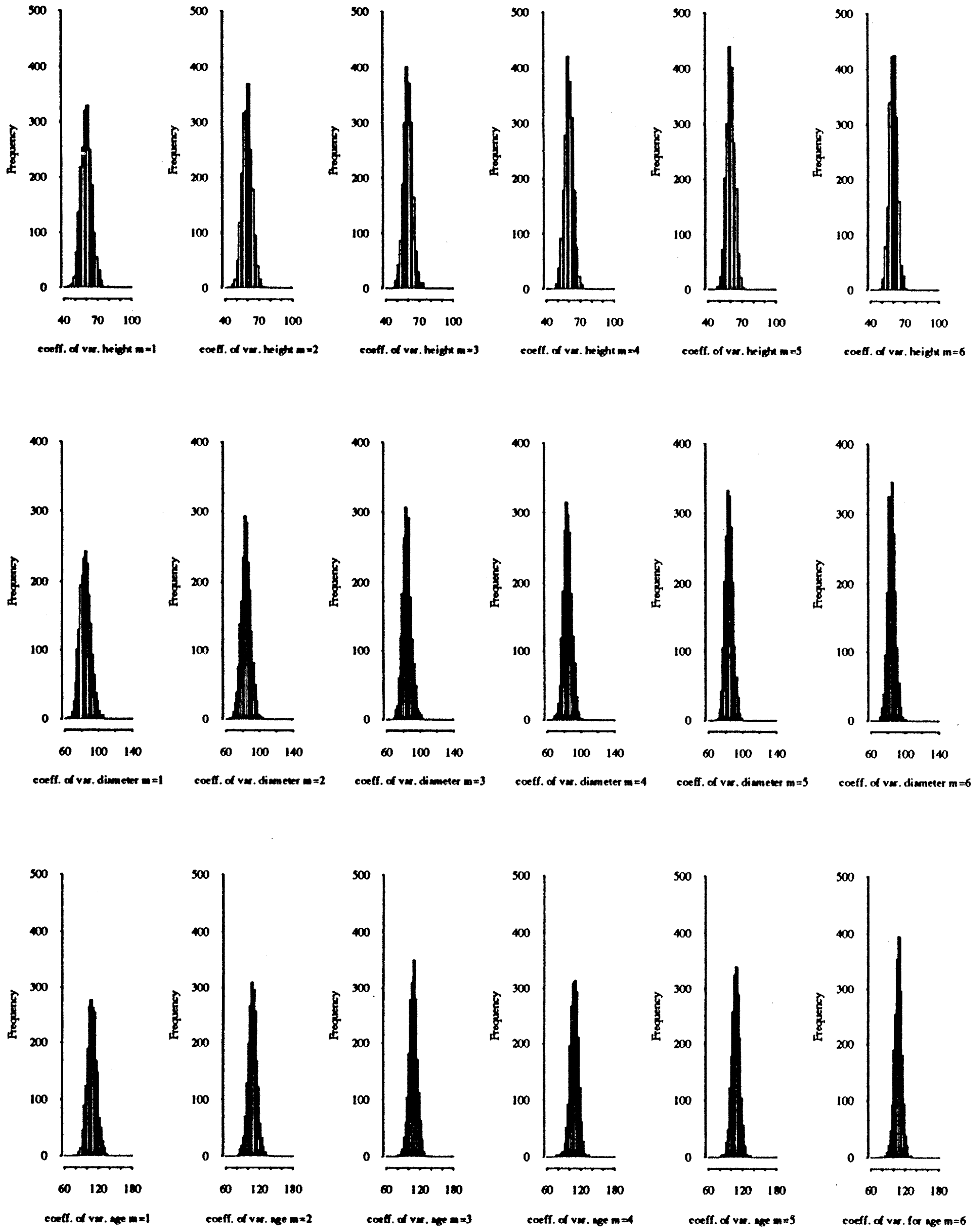

Figure 9. Sampling distribution of $\widehat{cv}(RSS)$ for height, diameter, and age while using the modified Takahasi's method.

other methods because of its superior performance. But if one thinks of using the collected RSS data in the future also for either comparison or combining (i.e., meta analysis) purposes without ranking information, or hopes to improve the ranking information with growing information base the modified Takahasi's RSS method needs to be considered.

References

Halls, L. K. and Dell, T. R. 1966. Trial of ranked set sampling for forage yields. Forest Science 12:22-26.

Johnson, G. D., Patil, G. P. & Sinha, A. K. 1993. Ranked set sampling for vegetation research. Abstracta Botanica 17:87-102.

McIntyre, G. A. (1952). A method for unbiased selective sampling, using ranked sets. Australian Journal of Agricultural Research 3:385-390.

Patil, G. P., Sinha, A. K., and Taillie, C. 1992a. Ranked set sampling: An annotated bibliography. Center for Statistical Ecology and Environmental Statistics, Pennsylvania State University, University Park, PA Technical Report Number 91-1201.

Patil, G. P., Sinha, A. K., and Taillie, C. 1992b. Ranked set sampling: A novel method to accomplish observational economy in environmental studies. Center for Statistical Ecology and Environmental Statistics, Pennsylvania State University, University Park, PA Technical Report Number 92-1004.

Patil, G. P., Sinha, A. K., and Taillie, C. 1993. Estimation of the population variance using ranked set samples. Center for Statistical Ecology and Environmental Statistics, Pennsylvania State University, University Park, PA Technical Report Number 93-0501

Patil, G. P., Sinha, A. K., and Taillie, C. 1994a. Ranked set sampling. In: G. P. Patil and C. R. Rao (eds.), Handbook of Statistics, Vol. 12: Environmental Statistics, North-Holland Elsevier, New York, 167-200.

Patil,G. P., Sinha, A. K., and Taillie, C. 1994b. Ranked set sampling for multiple characteristics. International Journal of Ecology and Environmental Sciences 20:94-109.

Platt, W. J., Evans, G. W., and Rathbun, S. L. 1988. The population dynamics of a long-lived conifer. The American Naturalist 131(4):491-525.

Rao, T. J. 1992. On certain problems of sampling designs and estimation for multiple characteristics. Technical Report Number 13/92, Stat-Math Unit, Indian Statistical Institute, 203 B. T. Road, Calcutta-700035, India.

Sarndal, C. E., Swensson, B., and Wretman, J. 1992. Model Assisted Sampling. Springer-Verlag, New York.

Takahasi, K. 1970. Practical note on estimation of population means based on samples stratified by means of ordering. Annals of the Institute of Statistical Mathematics 22:421-428.