Tree Planter's Notes, Volume 42, No. 4 (1991)

Bareroot Seedling Inventory: Estimation of Optimal Sample Size

Carl Mize and Dave Hansen

Associate professor and forestry student, Iowa State University, Ames, IA

Many State nurseries operate on an "inventory, sell, lift, and ship" schedule. One drawback of this schedule is that nurseries often end up with seedlings that cannot be sold, which results in reduced income. A fairly simple technique that allows nursery managers to estimate the number of plots needing to be measured during the inventory phase in order to achieve maximum profits is presented. The technique is discussed and an example is presented. Data supplied by seven State nurseries show consistent biases in estimated total number of salable seedlings. Tree Planters' Notes 42(4):9-13; 1991.

In recent years, rising costs, budget cuts, and competition from privately owned nurseries have made operational efficiency increasingly important for government forest nurseries. Because a significant part of the production of most nurseries is bareroot seedlings, managers need to look closely at this aspect of their operation to maximize profitability.

An important phase of the production process is estimating the number of seedlings that can be sold. We have studied how some nurseries estimate the number of salable seedlings and are presenting our findings in this paper. For nurseries selling seedlings based only on an estimate of the number of seedlings and not an actual count, we are also presenting a relatively simple technique to calculate how many plots need to be measured to estimate the number of seedlings.

Current Practices

In 1984, a questionnaire was sent to 56 State nurseries throughout the United States to determine existing practices and problems dealing with the inventory of bareroot seedlings. Thirty-one nurseries responded, and although management styles and customer bases varied widely, some generalizations could be made.

Most nurseries operated on an "inventory, sell, lift, and ship" schedule. The basic procedure is as follows:

?? Take an inventory to estimate the number of salable seedlings in the fall
?? Sell seedlings during the winter
?? Lift, package, and ship them in the spring.

The problem with this procedure is the uncertainty associated with selling seedlings based on an inventory estimate; the actual number of salable seedlings is not known until packaging is finished.

Nurseries inventoried each species and seedlot to obtain an estimate of the total number of salable seedlings. These inventories usually involved counting the number of seedlings meeting the nursery's standards in a number of 15-cm (6-inch) or 30-cm (12-inch)-wide plots (called frames) randomly or systematically spread throughout the seedbeds containing each species or seedlot. The number of plots measured varied widely among nurseries, but 20 plots were used fairly commonly, unless there were many beds, in which case more plots were used.

Because of normal seedling mortality and damage occurring during the lifting process, the number of seedlings that would be salable after lifting and packaging was expected to be less than the number estimated from the inventory. Moreover, because even if there was no mortality or damage, the estimate is not an actual count. Managers knew there was a 50-50 chance that the actual number would be less than the estimate. Because managers wanted to avoid overselling seedlings (which would require them to inform customers that they could not meet their commitment), they used a variety of techniques to adjust the estimated total. The most common technique was to reduce the estimated total by a percentage to arrive at a number that the manager felt could be sold safely.

Most managers indicated that the adjustments that they made were enough to ensure that they seldom ran out of seedlings. An unfortunate result of their success was that they almost always ended up with confirmed sales for fewer seedlings than the actual number of salable seedlings. Sometimes they could sell the extra seedlings, but other times they could not and had to destroy them, which reduced their income.

Another Method

Most of the adjustments that nursery managers made to their inventory estimates might be best described as educated guesses. None of the managers indicated that their adjustments were based on statistical techniques. We have developed a technique for adjusting the inventory total based on the level of confidence the manager wishes to have in the estimate. The technique should improve profitability for nurseries that cannot sell their extra seedlings, can be easily defended, and can be easily adapted to changing circumstances.

The technique is developed by starting with the formula for a **one-sided confidence interval.** For this paper, we define a one-sided confidence interval as a value (number of seedlings) for which the probability that the actual number of salable seedlings is greater than the value can be specified. For example, if an inventory for a species was made, a one-sided 95% confidence interval for the total was calculated, and a value of 150,000 was obtained, then 95 times out of 100, in situations like this, there would be at least 150,000 seedlings. In such a situation, only about 5% of the time will the nursery end up with less than 150,000 seedlings. Managers who wish to be more or less sure can increase or decrease the confidence level accordingly.

The formula for the one-sided confidence interval is as follows: number of salable seedlings = $T - t \bullet S_T$, (1)

where T = estimated number of salable seedlings from an inventory.

t = value from a t-table that depends on the sample size used to estimate T and the desired likelihood of not having enough seedlings to fill orders. (A t-value of 1.729 would be used for a one-sided 95% confidence interval for a sample of 20 frames, which has 19 degrees of freedom. The degrees of freedom is the sample size minus 1.).

S_T = standard error of the total derived from the sample used to estimate T. The formula for S_T is square root of: $[(N^2 * S^2)/n]$ where S^2 (the variance of the number of seedlings per plot) is calculated from the inventory, N is the population size (the length of the beds planted to the species or seedlot divided by the width of the plot), and n is the sample size (number of plots). The finite correction factor [(N-n)IN] is not used because the sample size, n, is usually small compared with N.

Both t and S_T are affected by the size of the sample used to estimate T, that is, the number of plots measured. As sample size increases, t and S_T decrease, and the value of the one-sided confidence interval becomes close to the actual population value. But measuring many plots is expensive. The challenge is to identify the sample size that represents an optimal balance between increased precision and increased cost of inventory.

Formula 2 shows a simple marginal return model that can be used to predict net income. It shows that net income (I) (not considering production costs, which are fixed and independent of the number of seedlings sold) is a function of t (the t-value); the value of individual seedlings (V); the standard deviation of the total (S_T); and the cost of measuring each sample plot (C).

$$I = (T - t \bullet S_T) \bullet V - n \bullet C \qquad (2)$$

By taking the derivative of formula 2, with respect to n (first substituting the formula used to calculate S_T), setting it equal to zero, and rearranging the terms, the optimum sample size is estimated by formula 3. The value of n must be solved in a stepwise manner (explained in the example below) because the value of t depends upon the value of n.

$$n = \left[\frac{t * V \bullet S \bullet N}{[2 \bullet C]}\right]^{2/3} \qquad 3)$$

To show how to solve for n, we will work an example using the following values, which are representative of the values provided in the questionnaires:

V	=	\$0.10 (10 cents per seedling)
S	=	19 (standard deviation of number of seedlings per plot)
N	=	*800* (length of bed divided by plot width)
C	=	\$1.50 (\$1.50 to sample a frame)

Aside from these values, we need a value of t to estimate n . But the value of t depends upon n and how confident the nursery manager wants to be of the nursery's not running out of seedlings. Because we do not known, we can just guess that n might be equal to 20. (Note: the initial guess is not very important, it will not affect the final answer.) Moreover, let's assume the manager wants to be 95% confident that the nursery will not run out of seedlings, so we will start with a t-value of 1.729, which is for a

95% one-sided confidence interval with 19 degrees of freedom. By using formula 3 and the values just described, we estimate that n equals 92. But that value was calculated using a t-value for a sample size of 20. For a sample size of about 92, the t-value would be smaller. Therefore, then is between 20 and 92. So we will guess that n = 80, which would have a t-value of 1.667. Using 1.667, n = 89, which would have almost the same t-value as 80. So the sample size that should be taken to maximize profit is 89. For practical purposes, n can be estimated usually within two or three steps. A person moderately familiar with using spread sheets should be able to develop a simple spread sheet that can estimate sample size.

An examination of a t-table shows that t-values change very slightly for sample sizes (n) above 30. Often, an inventory will require at least 30 plots. Therefore, a good initial estimate of the sample size would be achieved by just inserting the t value for infinite degrees of freedom (the one at the bottom of each column) in the equation and calculating n . Iterations are not really necessary if the estimate is 30 or more.

The relation between optimum sample size and projected net income (forgetting about fixed costs and assuming that only the number of seedlings estimated by the one-sided confidence limit are sold) can be quantified by adjusting formula 1 to include the cost of sampling and the value of the seedlings. The resulting formula is as follows:

$$\text{Income} = (T - t^*S_T)^* V - (n^*C) \qquad (4)$$

Formula 4 was used to examine the relation between income and the number of plots sampled, the variability of the number of seedlings per plot, and the cost of measuring a plot. Figure 1 illustrates the relation between income and sample size (number of plots measured), using the values previously listed, the appropriate t value, and calculating the S_T for each sample size.

If the nursery had sampled 89 plots (figure 1), the income, using formula 4, would have been $4,137. If the nursery were to use 20, which is commonly done, the income would be $4,000, or $137 less than the optimum. These calculations assume that only the number of seedlings estimated by the one-sided confidence interval are sold and do not consider production costs.

Much variability exists between nurseries and species and even between beds. Changing the values of the terms used in formula 4, however, does not alter the basic relation between sample size and income. For example, figures 2 and 3 illustrate the effect of different standard deviations (S) and sampling costs (C) for the example just worked out.

As seen in figures 2 and 3, the optimal sample size increases as the cost per plot decreases and the variability increases. The curves are fairly flat on top, which means that if the actual number of measured plots is "close" to the number of plots estimated by formula 2, income should be about the same. Income is reduced less by taking too many samples than by taking too few.

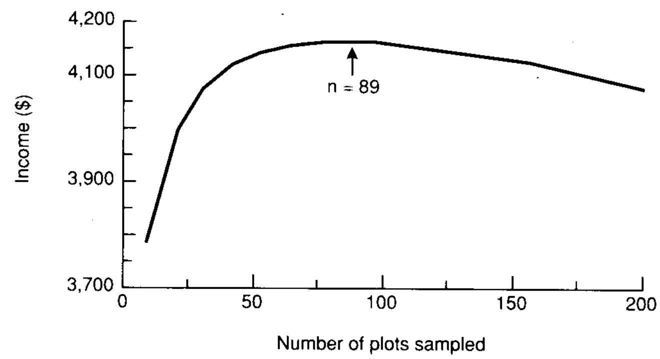

Figure 1—*Income versus sample size for hypothetical nursery bed.*

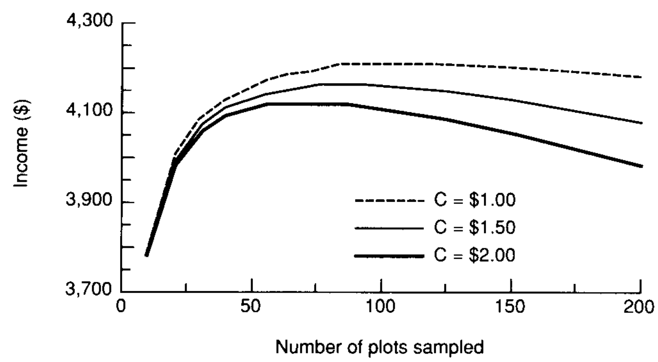

Figure 2—*Effect of different costs of measuring a plot on income.*

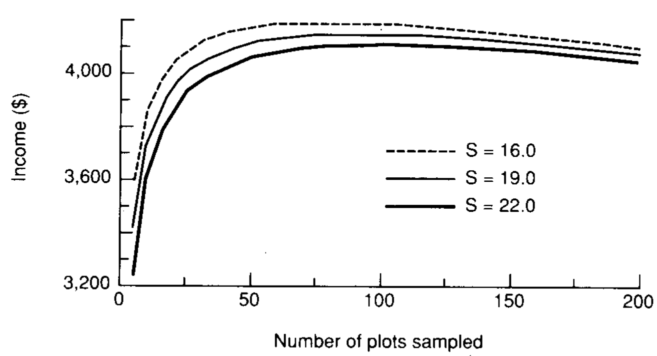

Figure 3—*Effect of different standard deviations on income.*

The value of seedlings is readily available to the nursery manager. The acceptable percentage of unfilled orders (tolerance level) will require managers to decide how often they want to contact purchasers and inform them that there are no seedlings available. Estimating the cost per plot will require an estimate of the time required to measure a plot and travel to another. (Note: some managers estimated that it cost nothing to do an individual plot. But assuming that it takes 2 workers, each making $9 per hour, 5 minutes to count a plot and travel to another plot, each plot will mean a cost of $1.50, so a cost of nothing is unrealistic.) The values of the standard deviation (S) can be taken from previous inventories if records are kept, or they can be calculated from samples of the seedbed being inventoried. (Note: if nurseries kept records of S for different species for some years, fairly consistent values may be observed.)

To make it easier to estimate the optimum sample size and to give nursery managers an indication of the sample sizes needed, we used equation 3 to develop a table of optimum sample sizes (table 1). Because equation 3 has five unknowns (t, V, S, N, and C), we decided to assume that V = $0.12 per plant, C = $1.50 per plot, and the likelihood used is 95%. This left two terms to vary, S and N. Values of S and *N* were chosen to span most of the range that nursery managers will likely encounter.

Table 1—*Optimum sample size for a range of standard deviations (5 to 30 SD) and bed lengths**

	No. of seedlings					
Bed length	5 SD	10 SD	15 SD	20 SD	25 SD	30 SD
400	45	71	92	112	130	147
800	71	120	147	178	206	233
2,000	130	206	270	327	380	429
4,000	206	327	429	520	603	681
10,000	380	603	790	957	1,111	1,254
20,000	603	957	1,254	1,519	1,763	1,991
50,000	1,111	1,763	2,310	2,799	3,247	3,667
75,000	1,455	2,310	3,037	3,667	4,255	4,805
100,000	1,763	2,799	3,667	4,442	5,155	5,821

*Using equation 3 and $t = 1.645$, $V = \$0.12$, and $C = \$1.50$.

Biased Estimates

For the system on sample size estimation to function, the estimate of the total, T, developed from the sample must be a good (unbiased) estimate. From the data provided by 7 nurseries, 6 nurseries showed a consistent tendency to underestimate their actual seedling populations. The seventh nursery tended to overestimate the actual number of seedlings. Thus, it seems likely that most nurseries produce biased estimates of the number of salable seedlings. For a nursery to decide whether or not the estimates of the total are biased requires records from previous inventories. An easy check would be to examine the ratio of estimated to actual number of seedlings and look for a tendency to over or under estimate the actual number of salable seedlings. If there is a tendency, the sampling procedure and the criteria for deciding whether a seedling is salable or not need to be reviewed.

Using Stratified Sampling

In many nurseries, the density of seedlings in beds seeded to a species can be quite variable. An important consequence of the variability is that the variance, S2, of the number of seedlings per sample plot will be higher than that in uniform beds. As a result, for a given number of plots, the one-sided confidence interval for the total number of seedlings in the variable beds will be lower than that for the uniform beds. One way of dealing with this is to use stratified sampling because it results in narrower confidence intervals for overall totals than do samples from a simple sample. If the manager can separate beds or portions of beds into different groups (strata) by density, then stratified sampling can be used effectively. A good reference on the use of stratified sampling is Freese (1962).

Summary and Conclusions

The optimum sample size for an inventory in a nursery bed is estimated by the following equation:

$$n = \frac{[t \bullet V \bullet S \bullet N]^{2/3}}{[2 \bullet C]}$$

where t is from a t-table, V is the value of an individual seedling, S is the standard deviation of the number of seedlings in a sample plot, *N* is the length of the bed(s) divided by the width of the sample plot, and C is the cost of measuring an individual sample plot.

Effective management of bareroot seedling stock can enhance the profitability of any nursery operation. Use of appropriate inventory procedures, including economically optimal sample sizes, is important to maximize effectiveness. Although time, personnel, and other considerations may influence the nursery manager to use less than the calculated optimal sample size, the technique discussed in this paper will provide the manager with additional information on the trade-offs of using various sample sizes.

Literature Cited

Freese, F. 1962. Elementary forest sampling. USDA Agriculture Handbook No. 232. 89 pp. [Reprinted by Oregon State University Book Stores, Corvalis, OR]

Acknowledgments

Journal paper J-14593 of the Iowa Agriculture and Home Economics Experiment Station, Ames; project 2667.

The authors thank the nurseries that returned their questionnaires and, particularly, those that sent inventory data. They also thank Paul Hinz, Statistics Department, Iowa State University, and Jerry Grebasch and Roger Jacob, Iowa Department of Natural Resources State Forest Nursery for their comments and suggestions concerning the preparation of this article.