FORESTRY
DEPARTMENTAL SERIES NO. 5
SEPTEMBER 1972

BASIC 3-P Sampling

Agricultural Experiment Station
AUBURN UNIVERSITY
R. Dennis Rouse, *Director* Auburn, Alabama

Basic 3-P Sampling

EVERT W. JOHNSON*

INTRODUCTION

SAMPLING with probability proportional to prediction, or 3-P Sampling, often is viewed as an esoteric procedure that has no application to the real world of forest inventory. This is unfortunate because the concept is far less difficult to master than is imagined and, more important, it can lead to very large reductions in inventory costs, particularly when used in conjunction with point-sampling and existing computer programs *(4,5,7,11)*. This publication was written in the hope of breaking down the barrier of misunderstanding surrounding the procedure and, perhaps, encouraging its general adoption by persons engaged in forest inventory operations.

In practice, 3-P sampling has been almost always associated with the use of dendrometers for upper stem measurements and with computer programs designed to process the large volumes of data that are generated by the procedure. The existing literature almost invariably incorporates these items into the discussion of 3-P sampling, perhaps tending to obscure the sampling procedure itself. In actuality, the sampling procedure is independent of both the dendrometer and computer and even has potential uses in many non-timber inventory situations. Because of this, dendrometry and computer programs will not be mentioned in this discussion except in the concluding section. The discussion will be as informal as feasible and will be carried out within the context of forest inventory. The procedure that forms the core of the discussion is a primitive form of 3-P sampling, which itself probably would have little general application. However, if a person understands this simple procedure he should have little difficulty understanding the more complex adaptations that are coming into use.

FRAME OF REFERENCE

Assume a partial cut is to be carried out and an estimate must be made of the volume or value of timber which has been marked for sale. Before the cut, each tree in the sale area must be visited, examined, judged, and then marked or left alone. This marking procedure yields no information on the volume marked or left. Some additional procedure is needed.

It would be possible to keep a "cut and leave" tally where a record is made of the number of trees in each category, usually by d.b.h. class. If the trees are large and valuable and the stand density is appropriately low, this procedure might be justified. However, in stands of small trees or trees of relatively low value, particularly where the stands are dense, the cost of even such a simple operation might be excessive.

Whenever the cost of a complete enumeration becomes excessive, the tendency is to look toward sampling for the needed information. One solution might be to divorce the data gathering part of the problem from the marking and to carry out an inventory using either fixed-radius plots or point-sampling after the marking is completed. Under appropriate statistical control, such an inventory probably would yield usable results and still be relatively inexpensive.

Another approach to the problem might be to sample the trees as they are marked. In this procedure, the ultimate sampling unit is the tree, not a piece of land surface which may or may not bear trees. Let us say that a cruise intensity of 10 per cent has been deemed appropriate in this case. This means that one of every 10 trees would be selected for sampling. The sampling could be systematic, in which every tenth tree would be measured, regardless of size or value. This approach, of course, would yield a valid estimate of average tree size, or value, provided bias were not present.[1] Bias could easily be introduced by the marker consciously or subconsciously choosing better-than-average (or worse-than-average) trees as the sample trees. This could have a serious effect on the accuracy of the inventory. A further problem with systematic sampling is that with it valid variability statistics cannot be obtained and such statistics are essential for valid interval estimates.

Use of random sampling would eliminate much of the problem of bias. Insofar as this inventory problem is concerned, the mechanics of random sampling would be simple and straightforward. The marker-cruiser could carry

[1] Bias occurs when the mean of all possible estimates of a given population's parameters (mean, total, variance, etc.), based on samples of a given size, does not equal the parameter. Where the population is not known in sufficient detail so that its component units can be listed so as to construct a frame, and if the sampling design is such that the sample size is indeterminate prior to the sampling, bias is apt to be present. These conditions exist in the problem used as the basis for this discussion, consequently, some bias may be present in all cases *(8,9)*. However, this bias will be ignored in the discussion since it probably would be small and, under the conditions of the problem, unavoidable.

* Professor, Department of Forestry.

a pouch on his belt in which there were 10 marbles, 9 white and 1 black. After each tree was marked, the cruiser would stir up the marbles in the pouch and make a random draw. If the marble were white the tree would not be measured and the man would move on. If the marble were black, however, the tree would be measured. In either case the marble would be returned to the pouch, making it ready for the next draw. Such a procedure would result in approximately 10 per cent of the marked trees being measured. The statistics of the inventory would be computed conventionally:[2]

$$\bar{x} = \frac{\sum_{i=1}^{n} X_i}{n}$$

where: $\bar{x}$ = mean tree volume or value
X_i = volume or value of the *ith* tree, and
n = number of trees measured;

$$s^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{x})^2}{n-1}$$

where: s^2 = variance of sample;

$$s = \pm \sqrt{s^2}$$

where: s = standard deviation of sample;

$$C.V. = s/\bar{x}$$

where: C.V. = coefficient of variation;

$$s^2_{\bar{x}} = \frac{s^2}{n} \left(1 - \frac{n}{N}\right)$$

where: $s^2_{\bar{x}}$ = variance of sample mean, and
N = total number of trees in population;

$$s_{\bar{x}} = \pm \sqrt{s^2_{\bar{x}}}$$

where: $s_{\bar{x}}$ = standard error of sample mean;

$$SE = \pm\, t\, s_{\bar{x}}$$

where: SE = sampling error for mean, and
t = Student's t value at the desired probability level and with n − 1 degrees of freedom;

$$SE_T = N(SE)$$

where: SE_T = sampling error for total;

$$T = N\bar{x}$$

where: T = total volume or value of timber marked for cutting.

To illustrate this, assume that on a certain tract 85 trees have been marked for cutting. Assume further that 9 of these trees were selected for measurement using the random sampling procedure described above. The resulting data are shown in Table 1 along with the computation of the estimate of the total value of the marked timber (\$476.93 ± \$245.07, at the 95% level of probability).

It is possible to obtain these same results using a somewhat different computation procedure. This, using the same data, is shown in Table 2. As can be seen, this is a more cumbersome and less efficient procedure than the one conventionally used and would rarely, if ever, be used in a real-life situation. However, the basic idea behind it is fundamental to the discussion of 3-P sampling and, consequently, must be understood.

TABLE 1. DATA AND COMPUTATIONS FROM A SIMPLE RANDOM SAMPLE USING CONVENTIONAL PROCEDURES

Sample tree no.	Measured tree value
	Dol.
1	11.50
2	4.50
3	2.00
4	7.50
5	3.00
6	12.50
7	2.50
8	3.00
9	4.00
Total	50.50

x = 50.50/9 = \$5.611 mean value per tree
T = 85 (\$5.611) = \$476.93 estimated total value of marked trees
s^2 = 15.7361
s = ± \$3.97 per tree
C.V. = 0.707 or 70.7%
$s_{\bar{x}}$ = ± \$1.25 per tree
SE, for trees, at the 95% probability level = ± \$2.88 per tree
SE, for total value, at the 95% probability level = 85(2.88) = ± \$245.07

TABLE 2. DATA AND COMPUTATIONS FROM A SIMPLE RANDOM SAMPLING USING THE MODIFIED COMPUTATION PROCEDURE

Sample tree no.	Weight	Measured tree value	Blow-up factor	Estimate of total value	
		Dol.		*Dol.*	*Dol.*
1	1	11.50	85/1=85	11.50 (85) =	977.50
2	1	4.50	85/1=85	4.50 (85) =	382.50
3	1	2.00	85/1=85	2.00 (85) =	170.00
4	1	7.50	85/1=85	7.50 (85) =	637.00
5	1	3.00	85/1=85	3.00 (85) =	255.00
6	1	12.50	85/1=85	12.50 (85) =	1,062.50
7	1	2.50	85/1=85	2.50 (85) =	212.50
8	1	3.00	85/1=85	3.00 (85) =	255.00
9	1	4.00	85/1=85	4.00 (85) =	340.00
Total	--	------	----------------	-----------	4,292.50

T = 4292.50/9= \$476.94 total value of marked trees
s^2 = 113,693.4028
s = ± \$337.18 per estimate of total
C.V. = 337.18/476.94 = 0.707 or 70.7%
s_x = ± \$106.27 per estimate of total
SE, for total, at 95% level of probability = ± \$245.07

Note the order of events in the two procedures. In the first, or conventional, method the mean value of the sample trees was computed first and then this was "blown-up" to an estimate of the total value by multiplying the mean value by the total count of trees in the population. In this case the blowing-up process came after the computation of the mean. In the second method, the value of each sample tree was first blown-up to an estimate of the total value and then these estimates of the total were averaged. The sequences of averaging and blowing-up are reversed. The results, however, are identical, within rounding.

[2] This is merely a brief review of elementary sampling theory. Most foresters learned this theory when they were students, but a few were not taught application of the theory.

Fundamental to the whole process, regardless of the procedure used, is the blow-up factor. It is the reciprocal of the probability of a given tree being chosen for measurement in a single random draw from the population.[3] In this case, the tree has 1 chance in 85 of being chosen if only one random draw is made, so the reciprocal of this probability, 85, is the blow-up factor. Thus, in the conventional procedure, the mean tree value is blown-up by 85 to arrive at the estimate of the total. In the second procedure, each sample tree value is independently blown-up by 85 to an estimate of the total, then these estimates are averaged. In the conventional procedure, the value of each sample tree is used as an estimate of the mean value. Each tree may have a value that is greater or smaller than the mean, but these estimates are exactly correct on the average across all trees in the population. Likewise, in the second procedure the estimates of the total value, obtained from the sample trees, may be individually too large or too small, but they are exactly correct on the average across all members of the population.

It should be noted that in the calculation of the mean in both procedures, each observation or data item has a weight of one. This fact of equal weights for the individual observations is characteristic of simple random sampling with equal probability.

This random sampling procedure is sound, unbiased, and practical, but it is inefficient. Each tree, regardless of size or value, has the same probability of being chosen. Consequently, trees of low value (small or defective trees, both with low volumes) are likely to be oversampled while highly valued trees are undersampled.

To overcome this problem, some method must be devised which would sample high-value trees more heavily than low-value trees. This, in short, would involve variable probability. Sampling with probability proportional to size (Bitterlich or point-sampling) would not be applicable since such an inventory could not be carried out until after the marking was completed. A further consideration is that value may or may not be closely correlated to size. For example, a 12-inch black walnut or black cherry would be worth a great deal more than a 12-inch post oak. Size also does not take into consideration defect and its relation to merchantability.

List sampling would not be appropriate since it could not be carried out until the marking was completed and a tentative value assigned to each tree. The trees would have to be marked with identifying numbers so that, if chosen for measurement from the list, they could be recovered. Finally, the field work of locating the chosen trees would be extremely laborious.

3-P SAMPLING

The idea of a tentative value placed on each tree can be used as the basis of a stratification scheme that can solve this inventory problem. A set of tree value classes or strata which are arbitrary but still representative of the tree population involved can be developed. To each of these classes a probability of being chosen is assigned.[4] These probabilities would be proportional to the class values so that high-value classes would have a relatively high probability and low-value classes would have a relatively low probability of being chosen. Each tree, following marking, would be assigned to one of these classes. Then, right on the spot, a random draw using the probability appropriate to that class would be made to see if the tree was to be measured.

Such a classification must be quick and easy to carry out and it must not require measurements, because much of the idea is to minimize measurement operations. The only logical procedure is to use ocular estimations of the value. Each tree would be assigned to a class using the marker's subjective judgment on size, quality, probable utility, local markets, and so forth. It is recognized that ocular estimating is subject to considerable random and systematic error when used for volume estimation. However, when it is used to assign a tree to a value class the only thing affected by such errors is the probability of the tree being chosen. If a low-value tree is erroneously assigned to a high-value class it merely means that particular tree has a greater chance of being sampled than it should. This would have little or no effect on the total volume estimates, but it would have an impact on the sampling error. This is explained later.

The procedure described above is 3-P sampling. Let us examine the idea within the context of an example. Assume that a series of four tree value classes has been established: (A) $10 per tree; (B) $8 per tree; (C) $6 per tree; and (D) $3 per tree. Assume further that the following probabilities of being chosen have been assigned to the classes: (A) 4 in 40; (B) 3 in 40; (C) 2 in 40; and (D) 1 in 40. A pouch containing 40 marbles is made up for each of the four classes. In the A class pouch, 4 marbles are black and 36 white. In the B class pouch, 3 are black and 37 white. In the C class pouch, 2 are black and 38 white, and in the D class pouch only 1 of the 40 marbles is black. The marker-cruiser carries all four of these pouches on his belt.

When the marker selects a tree, he classifies it ocularly by value class. He then stirs the marbles in the appropriate pouch and makes a random draw of one marble. If it is white, he records that a tree has been marked and denotes the class to which it is assigned, then he moves to the next tree. If the marble is black, the tree is measured and its size or value is recorded. The class to which it was assigned is also recorded. In either case, the marble is returned to the proper pouch so that all is ready for the next drawing. Notice that in this procedure the probability of a given tree being selected for measurement changes from class to class and a high-value tree has a greater chance of being measured than does a low-value tree. Thus, the important problem of preferential sampling has been solved.

COMPUTATIONS

Because of the varying probabilities used in this approach, the conventional computational procedures shown

[3] A further example of this principle might be a situation in which a 1,000-acre tract has been divided into 5,000 square, 1/5-acre, sample plots. Assume that one plot will be drawn at random from the 5,000 and the volume on it determined. Its probability of selection is 1/5,000. If one multiplies the volume on this plot by 5,000 (which is the reciprocal of 1/5,000) the product will be an estimate of the total volume. This estimate may be too high or too low but *the average of all possible such estimates* (5,000 of these) would be correct.

[4] The probabilities assigned are discussed at length in a following section of this publication.

earlier cannot be used. If 3-P data were used in such a computational procedure, the total volume or value would be strongly biased upward because proportionally more high-value than low-value trees would be sampled. However, the second procedure described earlier for simple random sampling can be used provided one change is made. For example, assume that the same 85 trees cruised by simple random sampling procedure were sampled according to the four classes described above. Further assume that the resulting data were as shown in Table 3. The computations are shown in Table 4. In Table 4 the blow-up factor is not constant across all trees. This is a consequence of the fact that the probabilities of the trees being chosen are not constant. A tree in class A has a probability of being chosen that is 4 times as great as does a tree in class D. In other words, the probabilities are weighted and the trees in class A have a weight of 4. The trees in class B have a weight of 3, those in class C have a weight of 2, while those in class D have a weight of 1.

TABLE 3. DATA FROM A 3-P SAMPLE

Class	Probability	Wt. units/tree	Trees assigned	Trees measured	Tree values obtained	Total weight
		No.	*No.*	*No.*	*Dol.*	
A	4:40	4	15	2	10.50, 11.50	4(15)=60
B	3:40	3	20	2	6.00, 8.00	3(20)=60
C	2:40	2	20	0		2(20)=40
D	1:40	1	30	1	2.00	1(30)=30
Total			85	5		190

TABLE 4. COMPUTATION OF 3-P SAMPLE

Sample tree no.	Wt.	Measured tree value	Blow-up factor	Estimate of total value	
		Dol.		*Dol.*	*Dol.*
1	4	10.50	190/4= 47.500	10.50(47.500)	= 498.75
2	4	11.50	190/4= 47.500	11.50(47.500)	= 546.25
3	3	6.00	190/3= 63.333	6.00(63.333)	= 380.00
4	3	8.00	190/3= 63.333	8.00(63.333)	= 506.67
5	1	2.00	190/1=190.000	2.00(190.000)	= 380.00
Total					2,311.67

T = 2311.67/5 = \$462.33 total value of marked trees
$s^2 = 5975.1343$
$s = \pm$ \$77.30 per estimate of total
C.V. = 16.7%
$s^2_{\bar{x}} = 1124.7593$
$s_{\bar{x}} = \pm$ \$33.54 per estimate of total
SE, for total, at 95% level of probability = ± \$93.10

To compute probabilities in a case such as this, it is necessary to sum the weights which were assigned and use this sum as the denominator in the probability fraction. In the case of Table 2, where the probabilities are constant across all the trees, each tree has a weight of 1 and the sum of the weights is 85. As a result, the probability of any individual tree being chosen in a single random draw is 1/85 and the blow-up factor is equal to 85. In the case of the 3-P sample shown in Table 3, 15 trees are assigned to class A. Each of these trees has a weight of 4 so the sum of weights for class A is 60. Correspondingly, the sum for class B is 60; for class C, 40; and for class D, 30. The grand total of these weights is 190, which is used to calculate the probabilities. Consequently the probability of a class A tree being chosen is 4/190, the probability of a class B tree is 3/190, and so forth. The reciprocals of these probabilities are the blow-up factors. When one considers the calculation procedures in Tables 2 and 4 in light of these ideas, it becomes obvious that, fundamentally, they are exactly alike. One simply has to use the proper probabilities, whether they are constant or variable.

All the terms used until the next section had been common before 3-P sampling was developed. No specialized 3-P sampling terms have yet appeared. For ready reference, terms are defined in the Appendix.

GETTING AWAY FROM THE POUCHES

In most cases, pouches of marbles would not be practical since many classes would be needed. The only truly practical solution is to use a set of random numbers. However, to set the stage for a description of how those random numbers should be set up and used, it appears desirable to describe a technique based on the use of a deck of cards. This also will provide a vehicle for introducing some of the special terminology and symbology associated with 3-P sampling.

The deck would be made up of cards bearing two kinds of labeling. First, there would be a card representing each of the classes which are to be recognized. Each of these cards would bear an integer. The magnitudes of these integers should be correlated with the volumes or values thought to be associated with the classes. For example, a deck that would yield results analagous to those obtained using the four pouches would have a card with a 4 (for class A), a 3 (for class B), a 2 (for class C), and a 1 (for class D). Note that these values correspond to the weights previously described. In 3-P terminology, there would be K cards bearing integers. In this case, K would be equal to 4.

All the remaining cards would be alike. They represent rejection of the tree and are analagous to the white marbles in the pouches. Therefore, they are called "rejection cards." These cards may be blank or they may bear symbols such as asterisks, black balls, zeroes, groups of X's, or whatever is desired. These rejection cards serve to control the probabilities of trees being chosen. There are Z such cards. Regardless of weight, the greater the value of Z the smaller is the probability of any tree being chosen and vice versa. Z, which can be any desired whole number, must equal at least 1 and probably should be greater than 1. More will be said about the magnitude of Z later in the discussion. In the case of the deck for the four pouch problem, Z would be equal to 36. This would bring the deck size up to 40, the same magnitude as the number of marbles in each pouch. The reasoning behind this will be made clear after the use of the deck has been explained.

The deck is used as follows. As before, each tree is visited, examined, judged, and either marked or not marked. If marked, it is then ocularly assigned to a class, and the class is recorded. Remember that now the class identification is in terms of the numbers mentioned in the paragraph about the K cards. In standard 3-P sampling terminology this class label is called a KPI number. The deck is now thoroughly shuffled and a card is drawn at random from the deck. If the card bears a rejection symbol, the tree is rejected and the cruiser moves on. If the card bears a number (called a KI value) that is larger than the class label (KPI value), the tree also is rejected. If, however, the card bears a number (KI value) that is equal to or smaller than the class label (KPI value), the tree is measured and evaluated. The resulting volume or value magnitude is called a YI value. Every tree in the forest has a YI value that is unknown until the tree is measured and evaluated.

How does probability enter into this procedure? To demonstrate this, let us compare the 40-card deck with the four pouches of marbles. Table 5 shows how the probabilities are determined by the deck. Note that these probabilities are identical with those associated with the pouches of marbles. An examination of this table also should reveal why the deck contains 36 rejection cards.

TABLE 5. HOW THE PROBABILITY OF A TREE BEING SAMPLED VARIES ACCORDING TO CLASS (KPI VALUE) USING THE 40 CARD DECK

Card	Action taken	Probability of Measurement	Probability of Rejection
When a tree is assigned to class (KPI)1:			
*	No measurement		36/40
4 > KPI	No measurement		1/40
3 > KPI	No measurement		1/40
2 > KPI	No measurement		1/40
1= KPI	Measure	1/40	
		1/40	39/40
	A tree in class (KPI)1 has 1 chance in 40 of being measured.		
When a tree is assigned to class (KPI)2:			
*	No measurement		36/40
4 > KPI	No measurement		1/40
3 > KPI	No measurement		1/40
2= KPI	Measure	1/40	
1 < KPI	Measure	1/40	
		2/40	38/40
	A tree in class (KPI)2 has 2 chances in 40 of being measured.		
When a tree is assigned to class (KPI)3:			
*	No measurement		36/40
4 > KPI	No measurement		1/40
3= KPI	Measure	1/40	
2 < KPI	Measure	1/40	
1 < KPI	Measure	1/40	
		3/40	37/40
	A tree in class (KPI)3 has 3 chances in 40 of being measured.		
When a tree is assigned to class (KPI)4:			
*	No measurement		36/40
4= KPI	Measure	1/40	
3 < KPI	Measure	1/40	
2 < KPI	Measure	1/40	
1 < KPI	Measure	1/40	
		4/40	36/40
	A tree in class (KPI)4 has 4 chances in 40 of being measured.		

From this point on the computations would be exactly the same as those shown in Table 4. In terms of the symbols used by Grosenbaugh *(1,2,7)*, the computations would take the following form:

K = number of classes and class cards = 4

Z = number of rejection cards = 36

m = number of trees in population = 85

KPI = class label, for class A = 4
for class B = 3
for class C = 2
for class D = 1

$\sum^{m}$ KPI = sum of weights = 190

$\dfrac{\sum^{m} \text{KPI}}{\text{KPI}}$ = blow-up factor, for class A = 190/4 = 47.500
for class B = 190/3 = 63.333
for class C = 190/2 = 95.000
for class D = 190/1 = 190.000

n = number of sample trees = 5

YI = value of the I*th* tree

$\text{TI} = \dfrac{\sum^{m} \text{KPI}}{\text{KPI}}(\text{YI})$ = estimate of the total volume or value on tract, based on the I*th* tree above

$\overline{\text{T}} = \dfrac{\sum^{n} \text{TI}}{n}$ = mean estimate of the total volume or value on the tract = \$462.33

PRECISION AND BIAS OF 3-P SAMPLES

In actual practice it nearly always would be desirable to use more than four classes. The actual number depends on several factors. To illustrate these factors, let us consider the 85-tree population of the four-class problem. Let us assume that the true values of the 15 marked trees in Class A average \$12.00 per tree, the 20 trees in Class B average \$8.00 per tree, the 20 trees in Class C average \$5.00 per tree, and the 30 trees in Class D average \$3.00 per tree. Under these circumstances the actual total value of the marked trees would be \$530.00. Now, instead of four classes, let us establish enough classes (KPI values) so that there would be a class (KPI value) for every possible tree value when those values are expressed in cents. This would require somewhat more than 1,500 classes (KPI values). If all 85 trees were correctly assigned to these classes, the sum of the weights would be:

$$\sum^{m} \text{KPI} = 15(1{,}200) + 20(800) + 20(500) + 30(300) = 53{,}000$$

The 15 in the first numerical term of the equation represents the 15 trees in old Class A and the 1,200 represents the average value in cents. The remaining terms are equivalent. The computations for the 5 tree sample used before would be as shown in Table 6.

Notice that in Table 6 each estimate of the total value is exactly correct. There is no variability. The variance and coefficient of variation are both equal to zero. Why did this occur? First, the classes were numerous enough for a class or KPI value to be assigned expressing a value in terms of cents, the smallest value unit. Second, these class (KPI) values were exactly the same as the actual tree (YI) values, therefore were perfectly correlated with them.

Such perfect correlation rarely exists. A more realistic situation is to assume that the class (KPI) values were in terms of dollars and that each tree had been correctly classified. Now there would be approximately 15 classes. The sum of the weights would be:

$$\sum^{m} KPI = 15(12) + 20(8) + 20(5) + 30(3) = 530$$

The computations are shown in Table 7. Notice that when the correlation between class (KPI) value and actual (YI) value is not perfect, as in trees 1 and 2, the individual estimates of the total value vary in magnitude, causing the variance and C.V. to take on values other than zero. Furthermore, as can be seen if one compares the C.V. values in Tables 4 and 7, the closer the KPI and YI values come to agreement the smaller the C.V. and, consequently, the greater the precision of the cruise. As a matter of fact, as can be seen in Table 8, the C.V. of the (YI/KPI) ratios is exactly equal to the C.V. of the estimates of the total. This all indicates that a designer of a 3-P sampling scheme should make every effort to keep the (YI/KPI) ratio as nearly constant as possible. The greater his success in accomplishing this the better will be the precision of the cruise.

Table 6. Computation of Total Value Using KPI Values Equivalent to Tree Values in Cents

Sample tree no.	Weight (or KPI value)	Measured tree value (YI)	Blow-up[1] factor	Estimates[2] of total value
		Dol.		*Dol.*
1	1,050	10.50	50.476	530.00
2	1,150	11.50	46.087	530.00
3	600	6.00	88.333	530.00
4	800	8.00	66.250	530.00
5	200	2.00	265.000	530.00
Total	------	------	------	2,650.00

$\bar{T} = 2{,}650.00/5 =$ \$530.00 total value of marked trees
$s^2 = 0$
$s = 0$
C.V. $= 0$
$s^2_{\bar{x}} = 0$
$s_{\bar{x}} = 0$
SE $= 0$

[1] e.g., 53,000/1,050 = 50.476, etc.
[2] e.g., \$10.50 (50.476) = \$530.00, etc.

Table 7. Computation of Total Value Using KPI Values Equilavent to Tree Values in Dollars

Sample tree no.	Weight (or KPI value)	Measured tree value (YI)	Blow-up[1] factor	Estimates[2] of total value
		Dol.		*Dol.*
1	11	10.50	48.182	505.91
2	11	11.50	48.182	554.09
3	6	6.00	88.333	530.00
4	8	8.00	66.250	530.00
5	2	2.00	265.000	530.00
Total	--	------	------	2,650.00

$\bar{T} = 2{,}650.00/5 =$ \$530.00 total volume of marked trees
$s^2 = 290.1641$
$s = \pm$ \$17.03 per estimate of total
C.V. = 0.032 or 3.2%
$s^2_{\bar{x}} = 54.6205$
$s_{\bar{x}} = \pm$ \$7.39 per estimate of total
SE, for total, at 95 % level of probability = ± \$20.52

[1] e.g., 530/11= 48.182, etc.
[2] e.g., \$10.50 (48.182) = \$505.91, etc.

[5] "Unbiased" means that on the average, over all possible attempts, the cruiser's mean assignment (KPI) value is equal to the true mean KPI value.

Table 8. Computation of the Coefficient of Variation of the (YI/KPI) Ratios When the KPI Values are Equivalent to Tree Values in Dollars

Ratio		
10.50/11 = 0.9545	$\bar{x}$ =	1.0000
11.50/11 = 1.0455	s^2 =	0.001033
6.00/6 = 1.0000	s =	± 0.0321
8.00/8 = 1.0000	CV =	0.0321 or 3.2%
2.00/2 = 1.0000		

A cruiser who is unbiased[5] but erratic in assigning trees to classes (KPI values) will make an unbiased estimate of the true total volume or value of the marked trees but his estimate will be low in precision because the (YI/KPI) ratios vary excessively.

When the cruiser is biased and consistently places trees in classes (KPI values) that are too low or too high, the estimate of the total may or may not be biased depending on the nature of the bias. If the cruiser's bias can be described as a ratio, either high or low, the resulting 3-P estimate of the total will be unbiased. For example, in the case of an unbiased cruiser an individual blow-up factor would be:

$$\frac{\sum^{m} KPI}{KPI}$$

In the case of a cruiser biased so as to consistently class trees a certain percentage too high or too low, an individual blow-up factor would be:

$$\frac{\sum^{m} (KPI)\cdot(Bias\%)}{(KPI)\cdot(Bias\%)},$$

which may be transformed to:

$$\frac{(Bias\%)\ (\sum^{m} KPI)}{(Bias\%)\ (\ \ KPI)},$$

in which case the bias terms cancel out and the usual blow-up factor would be left unchanged.

If the cruiser's bias is a constant number of classes (KPI values) too low or too high[6], the resulting estimate of the total will be biased. In such a case an individual blow-up factor would be:

$$\frac{\sum^{m} (KPI) + (Bias)}{(KPI) + (Bias)}.$$

There is no way by which the bias could be cancelled out or otherwise removed in this situation.

The effect of those two types of bias are demonstrated in Tables 9 and 10. The basic data in both cases are from the 85-tree population used previously. The entire population has been included in the calculations because only in this way can the bias be determined. The basic class (KPI) system used in these examples is that associated with dollar values. In Table 9 it is assumed that the cruiser consistently assigned class (KPI) values that were 10 per cent too high, while in Table 10 the assumption is

[6] This is also true if the bias is otherwise additive instead of multiplicative. This would occur whenever the discrepancies between YI and KPI values do not cancel out.

TABLE 9. COMPUTATION OF TOTAL VALUE WHEN CRUISER WAS CONSISTENTLY BIASED 10% HIGH ON CLASS ASSIGNMENTS

Old class	True weight (or KPI value)	Biased[1] weight	Trees	Measured value	Blow-up[2 3] factor	Estimate of total value[4]
			No.	Dol.		Dol.
A	12	13.2	15	12.00	44.167	7,950.06
B	8	8.8	20	8.00	66.250	10,600.00
C	5	5.5	20	5.00	106.000	10,600.00
D	3	3.3	30	3.00	176.667	15,900.03
Total	----	----	85	----	------	45,050.09

$\overline{T}$ = 45050.09/85 = $530.00 estimated total value of marked trees
530.00 true total value
$ 0.00 Bias in value

$s^2 = 0$[5]
s = 0
C.V. = 0.0

[1] e.g., 12 + 12(0.1) = 13.2, etc.
[2] $\sum^{m}$ KPI = 15(13.2) + 20(8.8) + 20(5.5) + 30(3.3) = 583
[3] e.g., 583/13.2 = 44.167, etc.
[4] e.g., 15 trees ($12.00)(44.167) = 7950.06, etc.
[5] Note that: $12.00(44.167) = $530.00
$ 8.00(66.250) = $530.00
$ 5.00(106.000) = $530.00
$ 3.00(176.667) = $530.00, there are no differences.

TABLE 10. COMPUTATION OF TOTAL VALUE WHEN CRUISER WAS CONSISTENTLY ONE CLASS HIGH ON CLASS ASSIGNMENTS

Old class	True weight (or KPI value)	Biased[1] weight	Trees	Measured value	Blow-up[2 3] factor	Estimate[4] of total value
			No.	Dol.		Dol.
A	12	13	15	12.00	47.308	8,515.44
B	8	9	20	8.00	68.333	10,933.28
C	5	6	20	5.00	102.500	10,250.00
D	3	4	30	3.00	153.750	13,837.50
Total	----	----	85	----	------	43,536.22

$\overline{T}$ = 43536.22/85 = $512.19 estimated total value of marked trees
530.00 true total value
−$ 17.81 Bias in value

s^2 = 1759.6662[5]
s = ± $41.95
C.V. = 0.0819 or 8.19%

[1] e.g., 12 + 1 = 13, etc.
[2] $\sum^{m}$ KPI = 15(13) + 20(9) + 20(6) + 30(4) = 615
[3] e.g., 615/13 = 47.308, etc.
[4] e.g., 15 trees ($12.00)(47.308) = 8515.44, etc.
[5] Note that: $12.00(47.308) = $567.70
$ 8.00(68.333) = $546.66
$ 5.00(102.500) = $512.50
$ 3.00(153.750) = $461.25

that the cruiser consistently assigned class (KPI) values that were one class too high. To save space all 85 trees are not listed. Instead the counts and mean values associated with the four original classes are used. This makes it necessary to use non-integer KPI values in Table 9, which should be acceptable since they are, in essence, means.

As can be seen in Table 9, a percentage bias has no effect on either the estimated total or the C.V. In spite of the fact that the class (KPI) values are not the same as the actual (YI) values, the (YI/KPI) ratio remains constant since the KPI values are expanded by a constant multiplier.

An additive bias, however, as shown in Table 10, changes both the estimated total and the C.V. In this case the (YI/KPI) ratio is changed because of the bias, resulting in considerable alterations in estimates of the total. It should be noted that the sign of the error in the estimated total is the opposite of the sign of the original cruiser's bias.

The evidence of Tables 9 and 10 indicates that persons charged with assigning class (KPI) values in the course of a 3-P inventory should be given training prior to doing the work and that every effort should be made during that training to break any tendency on the part of the cruiser to consistently assign classes a set number of units too high or too low.

Bias may occur in a 3-P sampling scheme from faulty design as well as from cruiser error. For example, assume that in the case of the situation shown in Table 7 the largest class (KPI value) was set at 12. This class would then be open-ended and any tree too large or too valuable to fall in class (KPI value) 11 would be assigned to class (KPI value) 12, even if it should have been assigned to a class with a larger KPI value. In this case the (YI/KPI) ratio would not be reasonably constant within class (KPI value) 12 and it would usually be higher (thus biased) than the ratios associated with the other classes. The effect can be seen in Table 11. It is assumed that two more of the 15 trees in original class A are included in the sample and their true (YI) values were $20 and $25, respectively. This is not reasonable under the original assumptions made about the population of trees since these trees deviate so strongly from the class mean of $12. However, this only exaggerates the bias and does not invalidate the idea being illustrated.

The upward bias under these circumstances is clearly evident. The effect of the open-ended class on the C.V.

TABLE 11. COMPUTATION OF TOTAL VALUE USING KPI VALUES EQUIVALENT TO TREE VALUES IN DOLLARS WITH AN OPEN-ENDED CLASS

Sample tree no.	Weight (or KPI value)	Measured tree value (YI)	Blow-up[1 2] factor	Estimate of total value[3]
		Dol.		Dol.
1	11	10.50	48.182	505.91
2	11	11.50	48.182	554.09
3	6	6.00	88.333	530.00
4	8	8.00	66.250	530.00
5	2	2.00	265.000	530.00
6	12	20.00	44.167	883.33
7	12	25.00	44.167	1,104.17
Total	----	----	------	4,637.50

$\overline{T}$ = 4637.50/7 = $662.50 estimated total value of marked trees
530.00 true total value
+$132.50 Bias

s^2 = 55463.3640
s = ± $235.51
C.V. = 0.3555 or 35.55%

[1] $\sum^{m}$ KPI = 530.
[2] e.g., 530/11 = 48.182, etc.
[3] e.g., $10.50(48.182) = $505.91, etc.

is also evident. These results emphasize that it would be better to have too many potential classes (KPI values), some of which would not be used, than it would be to have too few so that open-endedness occurs.

It would be possible to define a maximum class or KPI value and state prior to sampling that all the trees falling into this class would be measured (i.e., the sampling in this class would be 100 per cent). The trees included in this class would be completely excluded from the 3-P sample and would be considered as a separate part of the inventory problem. This procedure would eliminate the problem of an open-ended class.

CONTROL OF THE PROBABILITY OF SELECTION

In 3-P sampling the sampling intensity itself is subject to probability since no one knows prior to the cruise how many trees will be selected. If the same trees were cruised several times, the number of trees selected for measurement would vary.

Basically, the rate of sampling is dependent on two quantities: K, the number of classes (KPI values); and Z, the number of rejection cards in the deck. If K is held constant and Z is increased the likelihood of drawing a rejection card will increase. However, the magnitude of Z does not influence the relative probability of choosing a tree of a given class compared to the other classes (i.e., the relative weights of the classes remain constant). In terms of the original four classes, regardless of how large Z is, the trees in class A would be four times as likely to be selected for measurement as the trees in class D. In essence, Z acts as a dilutant.

The expected number of trees to be sampled (ESN) can be computed as follows:[7]

$$ESN = \frac{\sum^{m} KPI}{K + Z}$$

In any given situation the sum of weights ($\sum^{m}$ KPI) would not be known prior to the marking. Consequently, an estimate of the sum would have to be obtained from a light pre-sample. For this purpose a set of fixed-radius plots (or prism points) would be distributed over the area and each tree in each plot that is likely to be chosen for cutting at the time of marking would be assigned to a class (KPI value). There would be a sum of such KPI values from each plot. These sums would be averaged and put on a per acre basis, as is done with tree volumes. This per acre value would be multiplied by the number of acres to arrive at the estimate of $\sum^{m}$ KPI needed in the calculations.

K would be set by the inventory designer so that the desired range in classes (KPI values) would be available, preferably without any open-ended classes.

Z would then control ESN and determine the proportion of the population that would be considered for sampling. For example, using the original 4-class problem:

$$K = 4$$
$$Z = 36$$
$$\sum^{m} KPI = 190$$
$$ESN = 190/(4+36) = 4.75 \text{ or } 5 \text{ trees}$$

Had Z equalled 96:

$$ESN = 190/(4+96) = 1.90 \text{ or } 2 \text{ trees}$$

Had Z equalled 6:

$$ESN = 190/(4+6) = 19.0 \text{ or } 19 \text{ trees.}$$

To determine the magnitude of Z, it would be necessary to estimate the number of sample trees needed to achieve a specified level of precision at a specified probability level. This could be done by using standard formulae:[8]

$$n = \frac{t^2 s^2}{AE^2}$$

where: n = sample size
t = Student's t at the specified probability level and for the appropriate degrees of freedom
s^2 = variance of T values
AE = allowable sampling error in absolute terms

or

$$n = \frac{t^2 \, C.V.^2}{AE^2}$$

where: n = as above
t = as above
C.V. = coefficient of variation of T values
AE = allowable sampling error as a proportion of $\bar{T}$.

As an example, assume that an inventory is being designed with 12 classes (KPI values) so that K equals 12. Also assume that a pre-sample has been carried out and the total number of trees is estimated to be 85 and the sum of the weights ($\sum^{m}$ KPI) is estimated to be 530. The C.V. is found to be 0.03214. The allowable error is ± 2 per cent at the 95 per cent level of probability. Under these circumstances, n is found as follows, using iteration to reconcile n and the appropriate value of t:

First sample, t for ∞ d.f. = 1.960

$$n = \frac{t^2CV^2}{AE^2} = \frac{(1.960)^2 \, (0.03214)^2}{(0.02)^2} = 9.92 \text{ or } 10 \text{ trees}$$

Second estimate, t for 10-1 d.f. = 2.262

$$n = \frac{(2.262)^2 \, (0.03214)^2}{(0.02)^2} = 13.2 \text{ or } 14 \text{ trees}$$

---

[7] $ESN = \frac{\sum^{m} KPI}{K+Z} = \frac{KPI}{K+Z} + \frac{KP2}{K+Z} + \ldots + \frac{KPM}{K+Z}$

Each $\frac{KPI}{K+Z}$ is a probability (e.g., when KPI = 4, K = 4, and Z = 36, $\frac{KPI}{K+Z} = \frac{4}{40} = 0.1$).

In the case of the 85-tree population there would be 85 such ratios.

[8] These formulae are used since the populations usually are so large that corrections for finite populations would not be needed.

Third estimate, t for 14-1 d.f. = 2.160

$$n = \frac{(2.160)^2\,(0.03214)^2}{(0.02)^2} = 12.05 \text{ or } 13 \text{ trees}$$

Fourth estimate, t for 13-1 d.f. = 2.179

$$n = \frac{(2.179)^2\,(0.03214)^2}{(0.02)^2} = 12.26 \text{ or } 13 \text{ trees}$$

In this case, n stabilizes at 13 trees. To obtain Z, n is substituted for ESN in the ESN formulae:

$$n = \frac{\overset{m}{\Sigma}\ \text{KPI}}{K + Z}$$

and the equation is solved for Z

$$Z = \frac{\overset{m}{\Sigma}\ \text{KPI} - nK}{n}$$

Substituting the data from above:

$$Z = \frac{530 = 13(12)}{13} = 28.77 \text{ or } 29$$

To meet the inventory specifications, the deck of cards would contain 12 number cards and 29 rejection cards.

When designing a 3-P inventory there are two rules which must be observed concerning the relative magnitude of ESN (the expected size of sample). These are:

1) The product of ESN and the largest class (KPI) value should be less than the sum of weights ($\overset{m}{\Sigma}$ KPI). Applying this to the above example:

$$(\text{ESN})\ (\text{largest KPI}) = 13(12) = 156;\ \overset{m}{\Sigma}\ \text{KPI} = 530.$$

Hence, this requirement has been met.

2) The square of the ratio of Z to K should be greater than (4/ESN) − (4/m). e.g.,

$$(Z/K)^2 = (29/12)^2 = 5.84$$

$$(4/\text{ESN}) - (4/m) = (4/13) - (4/85) = 0.2606$$

$5.84 > 0.2606$, so this requirement also has been met.

WHAT HAPPENS WHEN NO TREES ARE SELECTED

It is possible in a cruise of this type that no trees at all would happen to be selected for measurement. This could lead to the erroneous conclusion that the volume or value was zero. Should this occur, the cruise must be repeated. If the provision is made in the inventory design that a repeat cruise would be made in the case of no tree's being selected, the probabilities of selection are changed. The corrections necessary to maintain exact probabilities are beyond the scope of this discussion. Reference is made to Grosenbaugh *(6)* and Space *(10)*.

If one wishes, the same tract could be cruised several times simultaneously by using interpenetrating samples. In such a case, each tree is assigned to a class (KPI value) in the normal manner, but instead of comparison with a single randomly drawn (KI) value this KPI value is then compared with several randomly drawn (KI) values. Each of these represents a separate cruise. The data of the several cruises must be kept separate and analyzed separately. This approach materially reduces the probability of ending up with the patently false estimate of zero volume or value. Again, the step-wise procedures are beyond the scope of this publication.

USING RANDOM NUMBERS RATHER THAN A DECK

A deck of cards such as that described earlier is not a practical tool in the woods. A practical substitute for the deck is a set of random numbers which vary in magnitude in the same manner as the numbered cards in the deck. Diluting these random numbers are rejection symbols which correspond to the rejection cards. These random numbers and the rejection symbols are generated by a computer using Grosenbaugh's RN3P program. The computer print-out from this program can be cut into strips, which are joined end-to-end and rolled on spools for use in a special random number dispenser that has been adequately described by Grosenbaugh *(2)* and Mesavage *(7)*.

WHERE DOES THE DENDROMETER FIT IN?

The 3-P sampling procedure is designed to select the trees that are to be measured or evaluated so that an efficient estimate of the total volume or value of the whole population can be made. Nowhere in the 3-P procedure are any requirements stated as to how the trees that are selected must be measured or evaluated. This evaluation procedure could consist of an ocular estimate; it could be based on a local, standard, or form-class volume table; or it could consist of a series of measurements made on the upper stem using the dendrometer. Though the dendrometer approach is preferred, it is not the only one that could be used.

The reason the dendrometer approach is preferred is that the evaluation of each tree is based on data from that tree alone. There is no recourse to volume tables, which had to be constructed from data obtained from other trees that might or might not have been similar to the trees being sampled. With the dendrometer procedure, measurement and estimation errors are minimized and considerably more accurate results can be obtained.

IN SUMMARY

The pure 3-P sampling design used as the foundation for this discussion is the simplest and probably the least useful that might be employed. A far more useful design involves the combination of 3-P sampling with point sampling which can be used for several types of inventories, including C.F.I. Grosenbaugh and his colleagues have developed a number of different designs. Grosenbaugh's STX computer program is sufficiently flexible to handle all of them. The combining of 3-P sampling with dendrometry and computer programming has made available to the forestry profession a powerful and versatile tool for inventory operations of all kinds. Foresters should take advantage of this tool whenever its use seems advantageous.

LITERATURE CITED

(1) GROSENBAUGH, L. R. 1963. Some Suggestions for Better Sample Tree Measurement. Proc. Soc. Amer. For., pp. 36-42.

(2) ——————. 1965. Three-pee Sampling Theory and Program 'THRP' for Computer Generation of Selection Criteria. USDA, Forest Service Research Paper PSW-21. 53 p.

(3) ——————. 1967. The Gains From Sample-Tree Selection with Unequal Probabilities. Forestry 65:203-206.

(4) ——————. 1967. STX-FORTRAN-4 Program for Estimates of Tree Populations from 3P Sample Tree Measurements. USDA, Forest Service Research Paper PSW-13 (Revised). 76 p.

(5) ——————. 1968. Sample Tree Measurement: A New Science. Forest Farmer, Dec., pp. 10-11.

(6) ——————. 1971. STX 1-11-71 for Dendrometry of Multistage 3P Samples. USDA, Forest Service. 63 p.

(7) MESAVAGE, C. 1971. STX Timber Estimating With 3-P Sampling and Dendrometry. USDA, Forest Service, Agr. Handbook No. 415.

(8) SCHREUDER, H. T., J. SEDRANSK, AND K. D. WARE. 1968. 3-P Sampling and Some Alternatives, I. Forest Science 14:429-454.

(9) —————————————— AND D. A. HAMILTON. 1971. 3-P Sampling and Some Alternatives, II. Forest Science 17:103-118.

(10) SPACE, J. C. 1971. Field Instructions, Three-P Forest Inventory. USDA, Forest Service, State & Private Forestry – SE Area. 18 p.

(11) STEBER, G. D. AND J. C. SPACE. 1972. New Inventory System Sweeping the South. J. Forestry 70: 76-79.

APPENDIX

Meanings of symbols conventionally included in the 3-P sampling literature.

K = The number of classes (or KPI values) to which trees can be assigned.

Z = The number of rejection cards in a 3-P sampling deck.

KPI = Class label. It is a whole number used to identify a class. This number is equal in magnitude to the weight assigned to the class.

KI = A value obtained from a random draw from the 3-P sampling deck provided an integer card is drawn.

YI = The actual value or volume of the tree in question. Every tree has a YI value but it is unknown until after the tree has been measured and evaluated.

m = The number of trees in the population being sampled.

n = The number of trees sampled.

$\sum^{m}$ KPI = The sum of the weights of all the trees in the population being sampled.

$\dfrac{\sum^{m} \text{KPI}}{\text{KPI}}$ = The blow-up factor for the I*th* tree.

TI = The estimate of the total value or volume in the population based on the I*th* tree alone.

$\overline{\text{T}}$ = The mean estimate of the total value or volume based on all the trees in the sample.

ESN = The expected number of sample trees.