United States
Department of
Agriculture

Forest Service

**Southern
Research Station**

Research Paper
SRS–34

Determining Sample Size for Tree Utilization Surveys

Stanley J. Zarnoch, James W. Bentley, and Tony G. Johnson

The Authors

Stanley J. Zarnoch is a Mathematical Statistician, U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC 28802; and **James W. Bentley** and **Tony G. Johnson** are Foresters, U.S. Department of Agriculture, Southern Research Station, Forest Service, Knoxville, TN 37919.

DISCLAIMER

October 2004

Southern Research Station
P.O. Box 2680
Asheville, NC 28802

Determining Sample Size for Tree Utilization Surveys

Stanley J. Zarnoch, James W. Bentley, and Tony G. Johnson

Abstract

The U.S. Department of Agriculture Forest Service has conducted many studies to determine what proportion of the timber harvested in the South is actually utilized. This paper describes the statistical methods used to determine required sample sizes for estimating utilization ratios for a required level of precision. The data used are those for 515 hardwood and 1,557 softwood trees harvested in east Texas and classified into 5 product types. Two-stage sampling was used to collect the utilization data. The primary units were the logging operation locations and the secondary units were the trees within locations. The ratio of means estimator was used to calculate each of three utilization ratios. However, for simplicity, the mean of ratios approach was used to develop the statistical methodology for estimating sample sizes for a specified level of precision, defined as half the width of the 95-percent confidence interval. The infinite population model was used and variance components for the two-stage nested analysis of variance were obtained using PROC MIXED. The three utilization ratios were computed for all product classes for hardwoods and softwoods, as were the standard errors and 95-percent confidence intervals. The variance components were then obtained and used to develop tables that yield sample size scenarios based on specified levels of precision.

Keywords: Mean of ratios, PROC MIXED, ratio of means, two-stage sampling.

Introduction

The U.S. Department of Agriculture Forest Service (Forest Service) has conducted studies to determine what proportion of the timber harvested in the South is actually utilized. Logging operations across a State are selected at random, a sample of the trees cut at each location is selected, and various utilization proportions are determined. A number of utilization ratios have been defined and are used in the calculation of these utilization proportions. Utilization ratios are ratios of volumes of tree sections used by the logger to volumes of whole trees or to volumes of various parts of trees.

This paper has three objectives. The first is to present the ratio of means estimator, standard error, and confidence intervals for each of three utilization ratios. The second is to describe the statistical methods for determining how many locations and trees per location must be sampled to obtain a desired error for the utilization ratios. The third is to provide tables that can be used to determine the sample size options for estimating the utilization ratios.

Methods

Utilization Ratios

The data used in this paper are taken from a harvest and utilization study that was performed in eastern Texas by the Forest Service, Forest Inventory and Analysis Research Work Unit (FIA) in 2003 (Bentley and Johnson 2004). The data were those for 515 hardwood and 1,557 softwood trees that were classified into 5 product types: saw log, veneer, composite panel, pulpwood, and pole. Tables 1, 2, and 3 show these classifications with the number of locations and trees per location that were actually sampled. At each location, one or more product types were sampled.

The three utilization ratios used in this study were defined based on the following variables that describe the types of sections obtained from each cut tree:

x_1 = saw-log sections utilized (FIA section codes 1 and 4)

x_2 = saw-log sections not utilized (FIA section codes 1 and 4)

x_3 = upper-stem sections utilized (FIA section codes 2 and 5)

x_4 = upper-stem sections not utilized (FIA section codes 2 and 5)

x_5 = stump section utilized (FIA section code 0)

x_6 = stump section not utilized (FIA section code 0)

x_7 = top and limb sections utilized (FIA sections 3, 6, 7, 8, and 9)

x_8 = top and limb sections not utilized (FIA sections 3, 6, 7, 8, and 9)

The three utilization ratios studied here are based on the total volume of the cut tree, volume of saw-log and upper-stem sections (total merchantable volume as defined by FIA) or the total used volume.

Table 1—Estimates of the R_1 utilization ratios, standard error, and 95-percent confidence intervals

Species group and product[a]	Locations sampled[b]	Trees actually sampled	R_1	Standard error	Lower	Upper
	- - - - - *number* - - - - -					
Hardwood						
Saw logs	18	179	0.757	0.0165	0.722	0.792
Veneer logs[c]	2	20	0.709	0.0002	0.706	0.711
Composite panels[c]	3	61	0.806	0.0196	0.721	0.890
Pulpwood	28	255	0.807	0.0124	0.781	0.832
All hardwoods	34	515	0.762	0.0143	0.733	0.791
Softwood						
Saw logs	45	581	0.865	0.0042	0.857	0.873
Veneer logs	26	417	0.865	0.0045	0.856	0.875
Composite panels[c]	5	91	0.879	0.0082	0.856	0.902
Pulpwood	37	438	0.883	0.0060	0.871	0.895
Poles[c]	1	30	0.844	—	—	—
All softwoods	68	1,557	0.866	0.0029	0.860	0.872

— = not calculated.

[a] There was no pole product for hardwood.

[b] Locations sampled do not total because more than one species group and/or product can be measured at an individual location; therefore, double counting of the sample location can occur.

[c] Not considered reliable because these estimates are based on very few sample locations.

Table 2—Estimates of the R_2 utilization ratios, standard error, and 95-percent confidence intervals

Species group and product[a]	Locations sampled[b]	Trees actually sampled	R_2	Standard error	Lower	Upper
	- - - - - *number* - - - - -					
Hardwood						
Saw logs	18	179	0.892	0.0129	0.865	0.920
Veneer logs[c]	2	20	0.880	0.0019	0.856	0.903
Composite panels[c]	3	61	0.941	0.0087	0.904	0.978
Pulpwood	28	255	0.950	0.0059	0.938	0.962
All hardwoods	34	515	0.903	0.0106	0.882	0.925
Softwood						
Saw logs	45	581	0.957	0.0038	0.949	0.964
Veneer logs	26	417	0.958	0.0036	0.951	0.966
Composite panels[c]	5	91	0.967	0.0199	0.912	1.022
Pulpwood	37	438	0.976	0.0048	0.967	0.986
Poles[c]	1	30	0.933	—	—	—
All softwoods	68	1,557	0.958	0.0027	0.953	0.963

— = not calculated.

[a] There was no pole product for hardwood.

[b] Locations sampled do not total because more than one species group and/or product can be measured at an individual location; therefore, double counting of the sample location can occur.

[c] Not considered reliable because these estimates are based on very few sample locations.

Table 3—Estimates of the R_3 utilization ratios, standard error, and 95-percent confidence intervals

Species group and product[a]	Locations sampled[b]	Trees actually sampled	R_3	Standard error	Lower	Upper
	- - - - - *number* - - - - -					
Hardwood						
Saw logs	18	179	0.977	0.0024	0.972	0.982
Veneer logs[c]	2	20	0.984	0.0011	0.970	0.998
Composite panels[c]	3	61	0.944	0.0167	0.872	1.016
Pulpwood	28	255	0.927	0.0082	0.910	0.944
All hardwoods	34	515	0.967	0.0044	0.958	0.976
Softwood						
Saw logs	45	581	0.981	0.0019	0.978	0.985
Veneer logs	26	417	0.979	0.0035	0.972	0.987
Composite panels[c]	5	91	0.889	0.0152	0.846	0.931
Pulpwood	37	438	0.914	0.0086	0.896	0.931
Poles[c]	1	30	0.981	—	—	—
All softwoods	68	1,557	0.974	0.0026	0.969	0.980

— = not calculated.

[a] There was no pole product for hardwood.

[b] Locations sampled do not total because more than one species group and/or product can be measured at an individual location; therefore, double counting of the sample location can occur.

[c] Not considered reliable because these estimates are based on very few sample locations.

Where the utilization ratio is defined as utilized volume over total volume of the cut tree, it may be expressed as

$$R_1 = \frac{x_1 + x_3 + x_5 + x_7}{x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8} \qquad (1)$$

Merchantable volume according to FIA standards is based only on saw-log and upper-stem sections. Where the utilization ratio is defined as FIA merchantable volume utilized over total FIA merchantable volume, it may be expressed as

$$R_2 = \frac{x_1 + x_3}{x_1 + x_2 + x_3 + x_4} \qquad (2)$$

Some logging operations, however, take timber that does not meet FIA merchantability standards. This consists of the stump and the top and limb sections. Where the utilization ratio is defined as FIA merchantable volume utilized over total volume used, it may be expressed as

$$R_3 = \frac{x_1 + x_3}{x_1 + x_3 + x_5 + x_7} \qquad (3)$$

Obviously, the complement of each of these ratios is itself another utilization ratio as defined by FIA, because it contains no new information it will not be discussed further.

Utilization Ratio Estimation

A two-stage sampling design (Cochran 1977) was used to collect the utilization data. The primary units are the logging operation locations and the secondary units are the trees within locations. The estimator for any of the utilization ratios is based on the ratio of means and is defined as

$$\hat{R} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} = \frac{\bar{y}}{\bar{x}} \qquad (4)$$

where

$y_i = \sum_{j=1}^{m_i} y_{ij}$ = the sum of the numerator component of the utilization ratio for location i

$x_i = \sum_{j=1}^{m_i} x_{ij}$ = the sum of the denominator component of the utilization ratio for location i

n = number of locations sampled
m_i = number of trees sampled in location i

The sample variance is

$$\hat{V}(\hat{R}) = \frac{1}{n\bar{x}^2}\left(S_y^2 + \hat{R}^2 S_x^2 - 2\hat{R}S_{yx}\right)$$
$$= \frac{1}{n(n-1)\bar{x}^2}\left(\sum_{i=1}^{n} y_i^2 + \hat{R}^2\sum_{i=1}^{n} x_i^2 - 2\hat{R}\sum_{i=1}^{n} y_i x_i\right) \quad (5)$$

where

S_y^2 and S_x^2 = the typical sample variances of the *y*s and *x*s, respectively
S_{yx}^2 = their covariance

The finite population correction has been ignored because the number of locations sampled is small with respect to the total number of locations in Texas. The typical standard error of $\hat{R}$ is the square root of $\hat{V}\left(\hat{R}\right)$, and confidence intervals at the $(1-\alpha)$ level are

$$\hat{R} \pm t_{(1-a),(n-1)}\sqrt{\hat{V}\left(\hat{R}\right)} \quad (6)$$

where

$t_{(1-a),(n-1)}$ = the two-tailed $(1-\alpha)$ value from the t-distribution with $(n-1)$ degrees of freedom

Two-Stage Sampling Design (Finite Population Model)

It is cumbersome to use the ratio of means estimator to calculate the reliability of the utilization ratios under alternative sample size scenarios for the two-stage survey design. Therefore, a mean of ratios approach based on an alternative estimator to equation (4) was employed to estimate the utilization ratios and their standard errors. Although the number of trees that FIA sampled per location varied, this statistical procedure will first be presented under the simplifying assumption of equal numbers and then generalized to the unequal numbers situation. The mean of ratios estimator for the utilization ratios assuming equal number of trees per location is

$$\hat{R}_{MOR} = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{y_{ij}}{nm} \quad (7)$$

where

y_{ij} = utilization ratio for tree j in location i
n = number of locations sampled
m = number of trees sampled in each location

The estimated variance is based on the finite population model and is defined as

$$\hat{V}\left(\hat{R}_{MOR}\right) = \left(1-\frac{n}{N}\right)\frac{s_1^2}{n} + \frac{n}{N}\left(1-\frac{m}{M}\right)\frac{s_2^2}{nm} \quad (8)$$

where

N = the total number of locations in the population
M = the total number of trees in each location

$$s_1^2 = \frac{\sum_{i=1}^{n}(\bar{y}_i - \bar{y})^2}{n-1} \quad (9)$$

and

$$s_2^2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij} - \bar{y}_i)^2}{n(m-1)} \quad (10)$$

Note that s_1^2 is commonly referred to as the between primary variance and that s_2^2 is the within primary variance.

The estimated utilization ratio actually used by FIA is not based on a mean of ratios approach as in equation (7). Instead, the ratio of means estimator [equation (4)] is used and the specific utilization ratio is defined as the average numerator value of all trees sampled divided by the average denominator value of all trees sampled. These estimators may lead to different results and the sample variance equations are also computed differently. However, the estimators are very similar in this study, and the standard errors are also. In addition, as stated previously, the sample variance for the mean of ratios approach is more appropriate for investigating sample size scenarios based on the number of locations and trees per location because it is a function of n and m. The ratio of means approach contains n but not m explicitly. Therefore, although the ratio of means is used for FIA utilization ratios (Bentley and Johnson 2004), the mean of ratios is used here to develop sample size scenarios.

Two-Stage Sampling (Infinite Population Model)

Although the utilization estimator and sample variance could be obtained by the finite population model approach, it is often easier and more informative to assume the infinite population model (Marcuse 1949). This is justifiable when n/N is negligible as is the case for the Texas survey, where the number of locations sampled is very small in relation to the total number of logging locations in Texas. Under the infinite population model the estimator is the same as in the finite population model [equation (7)] but the sample variance is now simply

$$\hat{V}\left(\hat{R}_{MOR}\right)=\frac{s_1^2}{n} \tag{11}$$

which is obtained from equation (8) by allowing n/N to approach 0. Furthermore, it is helpful to realize that this could be equated (after taking the expectation) to the typical variance components in a nested general linear model because

$$\frac{s_1^2}{n}=\frac{\sum_{i=1}^{n}(\bar{y}_i-\bar{y})^2}{n(n-1)}=\frac{m\sum_{i=1}^{n}(\bar{y}_i-\bar{y})^2}{mn(n-1)}$$
$$=\frac{MS_T}{mn}=\frac{\sigma_2^2+m\sigma_1^2}{mn}=\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{nm} \tag{12}$$

where the mean square for treatments in a nested analysis of variance is

$$MS_T=\frac{m\sum_{i=1}^{n}(\bar{y}_i-\bar{y})^2}{n-1} \tag{13}$$

and it's expected mean square is

$$E(MS_T)=\sigma_2^2+m\sigma_1^2 \tag{14}$$

where

σ_1^2 = the variance component for the primary units

σ_2^2 = the variance component for the secondary units

Thus, the infinite population model has estimated variance defined as

$$\hat{V}\left(\hat{R}_{MOR}\right)=\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{nm} \tag{15}$$

Equations (11) and (15) will yield the same estimated variance when the m_i s are equal. However, when the number of trees per location is not a constant, an estimate of the mean number of trees per location m_0 is used for m in equation (15) and is defined as

$$m_0=\frac{1}{n-1}\left(\sum_{i=1}^{n}m_i-\frac{\sum_{i=1}^{n}m_i^2}{\sum_{i=1}^{n}m_i}\right) \tag{16}$$

(Milliken and Johnson 1984, Montgomery 1976). The typical standard error of $\hat{R}_{MOR}$ is the square root of $\hat{V}\left(\hat{R}_{MOR}\right)$, and confidence intervals at the $(1-\alpha)$ level are

$$\hat{R}_{MOR}\pm t_{(1-a),(n-1)}\sqrt{\hat{V}\left(\hat{R}_{MOR}\right)} \tag{17}$$

where

$t_{(1-a),(n-1)}$ = the two-tailed $(1-\alpha)$ value from the t-distribution with $(n-1)$ degrees of freedom

The advantage of equations (15) and (16) is that they are applicable in the most general situation whether the m_i s are equal or not. Equation (11) is a function of the number of locations (n) and is useful for calculating the sample variance of a given survey. However, for designing a survey, equation (15) gives the sample variance under any alternative combination of locations and trees per location.

The variance components in equation (15) could be obtained by performing a nested analysis of variance using PROC GLM (SAS Institute Inc. 1989), equating the expected mean squares to their mean squares from the analysis and solving for the variance components. Such computations are not complex but may be tedious if many must be performed or if there are unequal trees per location, two situations that were present in this study. Another possibility is to use PROC VARCOMP (SAS Institute Inc. 1989), which computes the variance components directly; however, subsequent calculations must then be performed to get the estimated variance. To avoid these difficulties, PROC MIXED (SAS Institute Inc. 1989) can be used to fit a nested analysis of variance and to obtain an estimate of the ratio, variance components, and standard error directly. PROC MIXED is a relatively new procedure in SAS and is much more general than the older PROC GLM and PROC VARCOMP.

Specification of the Error

Information from the Texas survey could be used to determine the intensity of sampling that is required to achieve a specified error in future utilization studies. In the present study, the error is defined as half the width of the 95-percent confidence interval [equation (17)], specifically

$$Error = t_{(1-a),(n-1)}\sqrt{\hat{V}\left(\hat{R}_{MOR}\right)} \quad (18)$$

where the components are as defined in equations (15) and (17) for the infinite population model. This is used as an approximation for the error under the ratio of means approach, as previously discussed. The error is a function of the two variance components and the number of locations (n) and trees per location (m). Hence, given the estimated variance components from the Texas survey, the error could be determined for various combinations of n and m.

Results

Utilization Ratios

The three utilization ratios were computed for all product classes for hardwoods and softwoods using the ratio of means estimator equation (4) (tables 1, 2, and 3). The R_1 utilization ratio was 0.762 for all hardwoods combined and 0.866 for all softwoods combined, and indicates the amount of total tree volume (merchantable and nonmerchantable) that was utilized for these two species groups. Similar results were obtained for products within species groups, with utilization about 10 percent higher for softwoods than for hardwoods.

When only FIA-defined merchantable volume is considered, the R_2 utilization ratios were 0.903 for hardwoods and 0.958 for softwoods, which indicates that most of the merchantable volume was being utilized, especially for the softwoods. Similar trends were found for each product.

The R_3 utilization ratios differed little by species group, which shows that the FIA merchantable proportion of timber taken by the logger was about 0.967 for hardwoods and 0.974 for softwoods.

In addition to the utilization ratios, tables 1, 2, and 3 give the standard errors computed from equation (5) and 95-percent confidence intervals computed from equation (6). The width of these confidence intervals is a measure of the reliability of the utilization ratios and, thus, an indication of the adequacy of the initial survey.

Although these utilization ratios were computed for all products for both species groups, they are not considered reliable if they were based on very few sample locations. Thus, the hardwood veneer and composite panel and softwood composite panel and pole estimates should be viewed with caution.

Sample Size Options

To facilitate the calculation of sample sizes based on specified errors for future surveys, variance components were computed with PROC MIXED for both species groups and all products (table 4). The mean of ratios estimator was employed. Again, the hardwood veneer and composite panel and softwood composite panel and pole figures are given for completeness and should not be relied upon for future surveys. The sample size options are shown for R_1 in table 5, R_2 in table 6, and R_3 in table 7. These tables were constructed based on the error defined in equation (18) with a range of sampling locations (n) from 10 to 100 and a range of trees per location (m) from 10 to 30, which seems reasonable from an operational sampling perspective.

These tables can be used to find a sampling option that gives a specified error. For instance, if we are interested in the hardwood saw-log utilization ratio R_1 within 0.020 error we have to sample 50 locations and 10 trees per location (500 trees). This is found from table 5 by glancing down the hardwood saw-log column until an error of 0.020 or smaller is found. Often several alternatives are available with each alternative specifying a different number of locations and trees per location. In this example, a slightly smaller error (0.018) would require 60 locations and 10 trees per location (600 trees). Obviously, this slight decrease in the error is not worth the cost of sampling 10 more locations.

Generally, the error decreases most as the number of locations increases and decreases little with increases in the number of trees per location. For instance, consider the hardwood saw-log R_1 utilization ratio and assume that 1,000 trees will be sampled. It is found from table 5 that $n = 40$ locations with $m = 25$ trees per location gives an error of 0.021. However, if $n = 100$ locations with $m = 10$ trees per location, the error is only 0.014. It should be realized that the cost of traveling to 60 more locations has not been taken into account. The second option has a smaller error for the same number of trees, but it involves additional travel costs and may be inefficient for this reason.

Table 4—Estimates of the variance components for the utilization ratios computed with the mean of ratios approach using PROC MIXED

Species group and product	R_1		R_2		R_3	
	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$
Hardwood						
Saw logs	0.00426	0.00422	0.00428	0.00555	0.00004	0.00017
Veneer logs[a]	0.00000	0.00011	0.00000	0.00161	0.00002	0.00002
Composite panels[a]	0.00088	0.00655	0.00000	0.00445	0.00141	0.00684
Pulpwood	0.00373	0.00700	0.00002	0.00489	0.00255	0.00733
All hardwoods	0.00371	0.00612	0.00142	0.00514	0.00359	0.00473
Softwood						
Saw logs	0.00039	0.00200	0.00042	0.00237	0.00012	0.00011
Veneer logs	0.00020	0.00172	0.00017	0.00206	0.00023	0.00036
Composite panels[a]	0.00000	0.00723	0.00069	0.00948	0.00330	0.01570
Pulpwood	0.00049	0.00267	0.00005	0.00125	0.00252	0.00859
Poles[a]	0.00000	0.00332	0.00000	0.00386	0.00000	0.00001
All softwoods	0.00042	0.00249	0.00033	0.00246	0.00258	0.00410

[a] Not considered reliable because these estimates are based on very few sample locations.

These tables are presented for convenience, but it should be realized that the error could be computed from the statistical formulas previously given for any sample size option. As an example, to compute the hardwood saw-log R_1 utilization ratio where $n = 50$ and $m = 10$, obtain the variance components $\hat{\sigma}_1^2 = 0.00426$ and $\hat{\sigma}_2^2 = 0.00422$ from table 4. Then, using equations (15) and (18) and the t-distribution value at the 0.95 level with 49 degrees of freedom, the error is

$$Error = 2.00958\sqrt{\frac{0.00426}{50}+\frac{0.00422}{50(10)}} = 0.019$$

which is identical to that in table 5.

The error computed in tables 5, 6, and 7 is based on the 0.95 confidence level. However, if a lower or higher level is desired, the t-value that corresponds to this significance level would be used in equation (18). All other equations would remain the same and the error would be computed as explained in the above example.

When a survey is being designed, tables 5, 6, and 7 can be used to determine an appropriate sample size for the desired level of precision. However, the error and associated standard error is only an approximation that is based on the means of ratios approach. The utilization ratios should be obtained by the ratio of means method using equation (4) and the standard error and confidence interval from equations (5) and (6). All calculation should be based on only the new sample data.

Table 5—Error associated with sample size options for the R_1 utilization ratios

			Hardwoods			Softwoods			
n	m	Total	Saw logs	Pulpwood	All	Saw logs	Veneer logs	Pulpwood	All
10	10	100	0.049	0.048	0.047	0.017	0.014	0.020	0.019
10	15	150	0.048	0.046	0.046	0.016	0.013	0.018	0.017
10	20	200	0.048	0.046	0.045	0.016	0.012	0.018	0.017
10	25	250	0.048	0.045	0.045	0.016	0.012	0.017	0.016
10	30	300	0.047	0.045	0.045	0.015	0.011	0.017	0.016
20	10	200	0.032	0.031	0.031	0.011	0.009	0.013	0.012
20	15	300	0.032	0.030	0.030	0.011	0.008	0.012	0.011
20	20	400	0.031	0.030	0.030	0.010	0.008	0.012	0.011
20	25	500	0.031	0.030	0.029	0.010	0.008	0.011	0.011
20	30	600	0.031	0.029	0.029	0.010	0.008	0.011	0.010
30	10	300	0.026	0.025	0.025	0.009	0.007	0.010	0.010
30	15	450	0.025	0.024	0.024	0.009	0.007	0.010	0.009
30	20	600	0.025	0.024	0.024	0.008	0.006	0.009	0.009
30	25	750	0.025	0.024	0.023	0.008	0.006	0.009	0.009
30	30	900	0.025	0.024	0.023	0.008	0.006	0.009	0.008
40	10	400	0.022	0.021	0.021	0.008	0.006	0.009	0.008
40	15	600	0.022	0.021	0.021	0.007	0.006	0.008	0.008
40	20	800	0.021	0.020	0.020	0.007	0.005	0.008	0.007
40	25	1,000	0.021	0.020	0.020	0.007	0.005	0.008	0.007
40	30	1,200	0.021	0.020	0.020	0.007	0.005	0.008	0.007
50	10	500	0.019	0.019	0.019	0.007	0.005	0.008	0.007
50	15	750	0.019	0.018	0.018	0.007	0.005	0.007	0.007
50	20	1,000	0.019	0.018	0.018	0.006	0.005	0.007	0.007
50	25	1,250	0.019	0.018	0.018	0.006	0.005	0.007	0.006
50	30	1,500	0.019	0.018	0.018	0.006	0.005	0.007	0.006
60	10	600	0.018	0.017	0.017	0.006	0.005	0.007	0.007
60	15	900	0.017	0.017	0.017	0.006	0.005	0.007	0.006
60	20	1,200	0.017	0.017	0.016	0.006	0.004	0.006	0.006
60	25	1,500	0.017	0.016	0.016	0.006	0.004	0.006	0.006
60	30	1,800	0.017	0.016	0.016	0.006	0.004	0.006	0.006
70	10	700	0.016	0.016	0.016	0.006	0.005	0.007	0.006
70	15	1,050	0.016	0.015	0.015	0.005	0.004	0.006	0.006
70	20	1,400	0.016	0.015	0.015	0.005	0.004	0.006	0.006
70	25	1,750	0.016	0.015	0.015	0.005	0.004	0.006	0.005
70	30	2,100	0.016	0.015	0.015	0.005	0.004	0.006	0.005
80	10	800	0.015	0.015	0.015	0.005	0.004	0.006	0.006
80	15	1,200	0.015	0.014	0.014	0.005	0.004	0.006	0.005
80	20	1,600	0.015	0.014	0.014	0.005	0.004	0.006	0.005
80	25	2,000	0.015	0.014	0.014	0.005	0.004	0.005	0.005
80	30	2,400	0.015	0.014	0.014	0.005	0.004	0.005	0.005
90	10	900	0.014	0.014	0.014	0.005	0.004	0.006	0.005
90	15	1,350	0.014	0.014	0.013	0.005	0.004	0.005	0.005
90	20	1,800	0.014	0.013	0.013	0.005	0.004	0.005	0.005
90	25	2,250	0.014	0.013	0.013	0.005	0.003	0.005	0.005
90	30	2,700	0.014	0.013	0.013	0.004	0.003	0.005	0.005
100	10	1,000	0.014	0.013	0.013	0.005	0.004	0.005	0.005
100	15	1,500	0.013	0.013	0.013	0.005	0.004	0.005	0.005
100	20	2,000	0.013	0.013	0.013	0.004	0.003	0.005	0.005
100	25	2,500	0.013	0.013	0.012	0.004	0.003	0.005	0.005
100	30	3,000	0.013	0.012	0.012	0.004	0.003	0.005	0.004

Table 6—Error associated with sample size options for the R_2 utilization ratios

			Hardwoods			Softwoods			
n	m	Total	Saw logs	Pulpwood	All	Saw logs	Veneer logs	Pulpwood	All
10	10	100	0.050	0.016	0.031	0.018	0.014	0.009	0.017
10	15	150	0.049	0.013	0.030	0.017	0.013	0.008	0.016
10	20	200	0.048	0.012	0.029	0.017	0.012	0.008	0.015
10	25	250	0.048	0.011	0.029	0.016	0.011	0.007	0.015
10	30	300	0.048	0.010	0.029	0.016	0.011	0.007	0.015
20	10	200	0.033	0.011	0.021	0.012	0.009	0.006	0.011
20	15	300	0.032	0.009	0.020	0.011	0.008	0.005	0.010
20	20	400	0.032	0.008	0.019	0.011	0.008	0.005	0.010
20	25	500	0.031	0.007	0.019	0.011	0.007	0.005	0.010
20	30	600	0.031	0.006	0.019	0.010	0.007	0.005	0.009
30	10	300	0.026	0.008	0.016	0.010	0.007	0.005	0.009
30	15	450	0.025	0.007	0.016	0.009	0.007	0.004	0.008
30	20	600	0.025	0.006	0.015	0.009	0.006	0.004	0.008
30	25	750	0.025	0.005	0.015	0.008	0.006	0.004	0.008
30	30	900	0.025	0.005	0.015	0.008	0.006	0.004	0.008
40	10	400	0.022	0.007	0.014	0.008	0.006	0.004	0.008
40	15	600	0.022	0.006	0.013	0.008	0.006	0.004	0.007
40	20	800	0.022	0.005	0.013	0.007	0.005	0.003	0.007
40	25	1,000	0.021	0.005	0.013	0.007	0.005	0.003	0.007
40	30	1,200	0.021	0.004	0.013	0.007	0.005	0.003	0.006
50	10	500	0.020	0.006	0.012	0.007	0.006	0.004	0.007
50	15	750	0.019	0.005	0.012	0.007	0.005	0.003	0.006
50	20	1,000	0.019	0.005	0.012	0.007	0.005	0.003	0.006
50	25	1,250	0.019	0.004	0.011	0.006	0.005	0.003	0.006
50	30	1,500	0.019	0.004	0.011	0.006	0.004	0.003	0.006
60	10	600	0.018	0.006	0.011	0.007	0.005	0.003	0.006
60	15	900	0.018	0.005	0.011	0.006	0.005	0.003	0.006
60	20	1,200	0.017	0.004	0.011	0.006	0.004	0.003	0.005
60	25	1,500	0.017	0.004	0.010	0.006	0.004	0.003	0.005
60	30	1,800	0.017	0.003	0.010	0.006	0.004	0.002	0.005
70	10	700	0.017	0.005	0.010	0.006	0.005	0.003	0.006
70	15	1,050	0.016	0.004	0.010	0.006	0.004	0.003	0.005
70	20	1,400	0.016	0.004	0.010	0.006	0.004	0.003	0.005
70	25	1,750	0.016	0.004	0.010	0.005	0.004	0.002	0.005
70	30	2,100	0.016	0.003	0.010	0.005	0.004	0.002	0.005
80	10	800	0.015	0.005	0.010	0.006	0.004	0.003	0.005
80	15	1,200	0.015	0.004	0.009	0.005	0.004	0.003	0.005
80	20	1,600	0.015	0.004	0.009	0.005	0.004	0.002	0.005
80	25	2,000	0.015	0.003	0.009	0.005	0.004	0.002	0.005
80	30	2,400	0.015	0.003	0.009	0.005	0.003	0.002	0.005
90	10	900	0.015	0.005	0.009	0.005	0.004	0.003	0.005
90	15	1,350	0.014	0.004	0.009	0.005	0.004	0.002	0.005
90	20	1,800	0.014	0.003	0.009	0.005	0.003	0.002	0.004
90	25	2,250	0.014	0.003	0.008	0.005	0.003	0.002	0.004
90	30	2,700	0.014	0.003	0.008	0.005	0.003	0.002	0.004
100	10	1,000	0.014	0.004	0.009	0.005	0.004	0.003	0.005
100	15	1,500	0.014	0.004	0.008	0.005	0.003	0.002	0.004
100	20	2,000	0.013	0.003	0.008	0.005	0.003	0.002	0.004
100	25	2,500	0.013	0.003	0.008	0.005	0.003	0.002	0.004
100	30	3,000	0.013	0.003	0.008	0.004	0.003	0.002	0.004

Table 7—Error associated with sample size options for the R_3 utilization ratios

n	m	Total	Hardwoods Saw logs	Hardwoods Pulpwood	Hardwoods All	Softwoods Saw logs	Softwoods Veneer logs	Softwoods Pulpwood	Softwoods All
10	10	100	0.005	0.041	0.046	0.008	0.012	0.042	0.039
10	15	150	0.005	0.039	0.045	0.008	0.011	0.040	0.038
10	20	200	0.005	0.039	0.044	0.008	0.011	0.039	0.038
10	25	250	0.005	0.038	0.044	0.008	0.011	0.038	0.037
10	30	300	0.005	0.038	0.044	0.008	0.011	0.038	0.037
20	10	200	0.004	0.027	0.030	0.005	0.008	0.027	0.026
20	15	300	0.003	0.026	0.029	0.005	0.007	0.026	0.025
20	20	400	0.003	0.025	0.029	0.005	0.007	0.025	0.025
20	25	500	0.003	0.025	0.029	0.005	0.007	0.025	0.025
20	30	600	0.003	0.025	0.029	0.005	0.007	0.025	0.024
30	10	300	0.003	0.021	0.024	0.004	0.006	0.022	0.020
30	15	450	0.003	0.021	0.023	0.004	0.006	0.021	0.020
30	20	600	0.003	0.020	0.023	0.004	0.006	0.020	0.020
30	25	750	0.003	0.020	0.023	0.004	0.006	0.020	0.020
30	30	900	0.003	0.020	0.023	0.004	0.006	0.020	0.019
40	10	400	0.002	0.018	0.020	0.004	0.005	0.019	0.017
40	15	600	0.002	0.018	0.020	0.004	0.005	0.018	0.017
40	20	800	0.002	0.017	0.020	0.004	0.005	0.017	0.017
40	25	1,000	0.002	0.017	0.020	0.004	0.005	0.017	0.017
40	30	1,200	0.002	0.017	0.020	0.004	0.005	0.017	0.017
50	10	500	0.002	0.016	0.018	0.003	0.005	0.017	0.016
50	15	750	0.002	0.016	0.018	0.003	0.005	0.016	0.015
50	20	1,000	0.002	0.015	0.018	0.003	0.004	0.015	0.015
50	25	1,250	0.002	0.015	0.017	0.003	0.004	0.015	0.015
50	30	1,500	0.002	0.015	0.017	0.003	0.004	0.015	0.015
60	10	600	0.002	0.015	0.016	0.003	0.004	0.015	0.014
60	15	900	0.002	0.014	0.016	0.003	0.004	0.014	0.014
60	20	1,200	0.002	0.014	0.016	0.003	0.004	0.014	0.014
60	25	1,500	0.002	0.014	0.016	0.003	0.004	0.014	0.014
60	30	1,800	0.002	0.014	0.016	0.003	0.004	0.014	0.013
70	10	700	0.002	0.014	0.015	0.003	0.004	0.014	0.013
70	15	1,050	0.002	0.013	0.015	0.003	0.004	0.013	0.013
70	20	1,400	0.002	0.013	0.015	0.003	0.004	0.013	0.013
70	25	1,750	0.002	0.013	0.015	0.003	0.004	0.013	0.012
70	30	2,100	0.002	0.013	0.015	0.003	0.004	0.013	0.012
80	10	800	0.002	0.013	0.014	0.003	0.004	0.013	0.012
80	15	1,200	0.002	0.012	0.014	0.003	0.004	0.012	0.012
80	20	1,600	0.002	0.012	0.014	0.002	0.004	0.012	0.012
80	25	2,000	0.002	0.012	0.014	0.002	0.003	0.012	0.012
80	30	2,400	0.002	0.012	0.014	0.002	0.003	0.012	0.012
90	10	900	0.002	0.012	0.013	0.002	0.003	0.012	0.011
90	15	1,350	0.002	0.012	0.013	0.002	0.003	0.012	0.011
90	20	1,800	0.001	0.011	0.013	0.002	0.003	0.011	0.011
90	25	2,250	0.001	0.011	0.013	0.002	0.003	0.011	0.011
90	30	2,700	0.001	0.011	0.013	0.002	0.003	0.011	0.011
100	10	1,000	0.001	0.011	0.013	0.002	0.003	0.012	0.011
100	15	1,500	0.001	0.011	0.012	0.002	0.003	0.011	0.011
100	20	2,000	0.001	0.011	0.012	0.002	0.003	0.011	0.010
100	25	2,500	0.001	0.011	0.012	0.002	0.003	0.011	0.010
100	30	3,000	0.001	0.010	0.012	0.002	0.003	0.011	0.010

Literature Cited

Bentley, J.W.; Johnson, T.G. 2004. Eastern Texas harvest and utilization study, 2003. Resour. Bull. SRS–97. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 28 p.

Cochran, W.G. 1977. Sampling techniques. 3rd ed. New York: John Wiley. 428 p.

Marcuse, S. 1949. Optimum allocation and variance components in nested sampling with an application to chemical analysis. Biometrics. 5: 189–206.

Milliken, G.A.; Johnson, D.E. 1984. Analysis of messy data. Volume I: designed experiments. New York: Van Nostrand Reinhold Co. 473 p.

Montgomery, D.C. 1976. Design and analysis of experiments. New York: John Wiley. 418 p.

SAS Institute Inc. 1989. SAS/STAT user's guide. Version 6. 4th ed. Cary, NC: SAS Institute Inc. 846 p. Vol. 2.

Zarnoch, S.J.; Bentley, J.W.; Johnson, T.G. 2004. Determining sample size for tree utilization surveys. Res. Pap. SRS–34. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 11 p.

The U.S. Department of Agriculture Forest Service has conducted many studies to determine what proportion of the timber harvested in the South is actually utilized. This paper describes the statistical methods used to determine required sample sizes for estimating utilization ratios for a required level of precision. The data used are those for 515 hardwood and 1,557 softwood trees harvested in east Texas and classified into 5 product types. Two-stage sampling was used to collect the utilization data. The primary units were the logging operation locations and the secondary units were the trees within locations. The ratio of means estimator was used to calculate each of three utilization ratios. However, for simplicity, the mean of ratios approach was used to develop the statistical methodology for estimating sample sizes for a specified level of precision, defined as half the width of the 95-percent confidence interval. The infinite population model was used and variance components for the two-stage nested analysis of variance were obtained using PROC MIXED. The three utilization ratios were computed for all product classes for hardwoods and softwoods, as were the standard errors and 95-percent confidence intervals. The variance components were then obtained and used to develop tables that yield sample size scenarios based on specified levels of precision.

Keywords: Mean of ratios, PROC MIXED, ratio of means, two-stage sampling.