

The Best Number of Trees to Measure per Plot Ignores the Budget, and Doubling or Halving It Costs the Same

— the best number is set by the ratio of costs and the ratio of variances, and the loss from missing it is set only by the share of the plot cost taken by the trees' total cost and the factor missed —the separator is whether the variation lies between plots or within them — [Paper 980]

Yuuki YAMAGISHI (山岸 勇輝)

Yamagishi Complexity Torus Lab, Japan

September 2026

YOI: <https://www.yamagishiyuuki.com/yoi/980>

— *From thinking to syncing.* —

Abstract

Stratum boundaries at equal steps of cumulative root frequency come close to optimal, while equal counts do worst (Paper 957). Subsampling within plots and differences in measuring cost were not treated there. Splitting the timber-volume distribution of Paper 957 (log-normal with log mean 4.5 and log standard deviation 0.8) into a plot effect and a tree effect, this paper takes n plots and m trees per plot and finds, by formula, exhaustive search and random sampling, the m that minimises the variance of the mean under a cost per plot and a cost per tree. No new theorem or law is claimed. First, the best number does not depend on the budget: with plot-effect share 0.1 and plot cost 20, the formula gives **15.8518** trees, and exhaustive search gave 16 trees for budgets of 1000, 2000 and 5000 alike. In all four combinations the best was the same for the three budgets (Section 2). Second, measuring twice or half as many costs the same: at twice and at half the best number the variance was the same **1.123327** times the best (share 0.1, cost 20), and 1.064604 to 1.123327 times across the four combinations (Section 3). Third, the size of the loss is set by one share: with α the share of the plot cost taken by the trees' total cost at the best, the loss agreed within 10^{-12} with a formula in α and the factor missed alone. Measuring one tree gave **4.432163** times and measuring all 50 gave **1.362887** times (share 0.1, cost 20) (Section 3). Fourth, random sampling follows the formula: the variance of the mean over ten thousand fresh draws from the population was 0.983340 to **1.033172** times the formula in six settings (Section 4). Fifth, the separator is whether the variation lies between plots or within them: raising the plot-effect share from 0.1 to 0.3 lowered the best number at cost 20 from 15.8518 to **8.0440** trees. The sum of the two variances stayed at **13776.529821** (Section 5). **Keywords:** two-stage sampling, survey sampling, optimal subsample size, survey cost, variance decomposition, forest inventory.

Scope of this paper (scope note): No new theorem or law is claimed: that in two-stage sampling the variance of the mean is the between-plot variance over the number of plots plus the within-plot variance over the total number of trees, and that under costs the best number is the square root of the product of the cost ratio and the variance ratio and does not depend on the budget, are known in sampling theory. These are set values: the volume distribution is the same log-normal as Paper 957; plot-effect shares of 0.1 and 0.3, plot costs of 20 and 100, a tree cost of 1 and budgets of 1000, 2000 and 5000 are set here. They correspond to no forest or species. Sampling is simplified: the population is taken as infinite; finite-population corrections, combination with strata, costs depending on travel distance between plots, and varying the number per plot are not included. No measuring plan is recommended: how many trees a real survey should measure is not treated. Relation to earlier papers: Paper 938, in its scope note, did not treat two-stage sampling that further subsamples within plots, and Paper 957, in its scope note, also did not treat subsampling within plots or different measuring costs by stratum; this paper sits in the seat of subsampling within plots (no earlier occupant; Paper 957 sat in the stratum-boundary seat of the same scope note of Paper 938). Paper 957 showed that stratum boundaries at equal steps of cumulative root frequency come close to optimal while equal counts do worst. Here, with the same distribution, the relation between how many trees to measure per plot and cost is counted. What is added: finding the best number by formula and exhaustive search for four combinations of costs and variances and comparing three budgets, confirming that the loss from missing it matches a formula in one share, comparing random sampling with the formula, seeing how far the best number moves when the variances are estimated from one forest, and placing the separator on “variation between plots, or within”.

1. Notation and the Question

Setting: tree volume is $V = \exp(4.5 + b + e)$, with plot effect b and tree effect e normal with mean 0 and variances $0.64s$ and $0.64(1 - s)$ (s is the plot-effect share, 0.1 or 0.3). The variance of the true plot means (between-plot variance) and the mean within-plot variance are

$$S_b^2 = e^{9+0.64(1-s)} e^{0.64s} (e^{0.64s} - 1), \quad S_w^2 = e^{9+1.28s} e^{0.64(1-s)} (e^{0.64(1-s)} - 1)$$

With n plots and m trees per plot, the variance of the mean and the number of plots under budget B , plot cost c_1 and tree cost c_2 are

$$\text{Var}(m) = \frac{S_b^2}{n} + \frac{S_w^2}{nm}, \quad n = \frac{B}{c_1 + m c_2}, \quad m^* = \sqrt{\frac{c_1 S_w^2}{c_2 S_b^2}}$$

where m^* minimises the variance. $c_1 = 20$ and 100 , $c_2 = 1$, and exhaustive search covered $m = 1$ to 200 .

Question: how many trees per plot is best, how does that change with the budget, and how much is lost by missing it?

2. The Best Number

Computation 2.1: true variances and the best number.

share s	between-plot variance	within-plot variance	plot cost	formula best	exhaustive (budgets 1000, 2000, 5000)
0.1	1015.664610	12760.865212	20	15.8518	16, 16, 16
0.1	1015.664610	12760.865212	100	35.4458	35, 35, 35
0.3	3252.813592	10523.716229	20	8.0440	8, 8, 8
0.3	3252.813592	10523.716229	100	17.9868	18, 18, 18

The ratio of within- to between-plot variance was 12.564054 for share 0.1 and 3.235266 for 0.3. In all four combinations the exhaustive best was the same for the three budgets and was the formula's best rounded down or up.

Observation 2.2

the best number does not depend on the budget. $\text{Var}(m)$ can be written $(S_b^2 + S_w^2/m)(c_1 + mc_2)/B$, and the budget B only divides the whole: raising the plot cost from 20 to 100 raised the best number for share 0.1 from 15.8518 to 35.4458 trees.

3. The Loss from Missing It

Computation 3.1: dividing the variance at k times the best number by the best variance gives, with $a = m^*c_2/c_1$,

$$\frac{\text{Var}(km^*)}{\text{Var}(m^*)} = \frac{1 + a(k + 1/k) + a^2}{(1 + a)^2}$$

which takes the same value for k and $1/k$.

share s	plot cost	a	twice	half	one tree	all 50
0.1	20	0.792592	1.123327	1.123327	4.432163	1.362887
0.1	100	0.354458	1.096606	1.096606	7.467578	1.023093
0.3	20	0.402198	1.102280	1.102280	2.261785	1.895303
0.3	100	0.179868	1.064604	1.064604	3.072804	1.147238

In all four combinations the values for twice and half agreed with each other and with the formula within 10^{-12} .

Observation 3.2

the loss is symmetric in the logarithm of the factor by which the number is missed. The larger a , the larger the loss for the same factor (at twice, from 1.064604 for $a = 0.179868$ to 1.123327 for $a = 0.792592$): the losses from measuring one tree and from measuring all 50 are set by how far each is from the best as a factor and by a ; for share 0.1 and cost 100, one tree gave 7.467578 and all 50 gave 1.023093.

4. Checking with Random Sampling

Computation 4.1: with plot cost 20 and budget 2000, the number of plots was $n = \lfloor 2000/(20 + m) \rfloor$, and plot and tree effects were drawn fresh from the population to form the mean ten thousand times.

share s	trees per plot	plots	sampled variance	formula	ratio
0.1	1	95	148.239353	145.016103	1.022227
0.1	16	55	32.920832	32.967612	0.998581
0.1	50	28	45.067825	45.388640	0.992932
0.3	1	95	142.843287	145.016103	0.985017
0.3	8	71	66.476279	64.341945	1.033172
0.3	50	28	121.628233	123.688854	0.983340

In six settings the sampled variance was 0.983340 to 1.033172 times the formula. For both shares the variance at the best number was smaller than with one tree and with all 50.

5. The Separator

Observation 5.1

the separator is whether the variation lies between plots or within them. Raising the plot-effect share from 0.1 to 0.3 lowered the best number from 15.8518 to 8.0440 trees at cost 20 and from 35.4458 to 17.9868 at cost 100: the sum of the two variances was 13776.529821 for both shares, and the formula variance with one tree was the same 145.016103 for both. What changed was only how the variation divides between and within plots.

Observation 5.2

estimated from one forest, the best number moves. Estimating the variances from one forest of 400 plots with 50 trees each gave, for share 0.1, a between-plot variance of 925.245704 and a within-plot variance of 12609.537335, and a best of 16.5096 trees at cost 20 (exhaustive 17): one tree away from the true 15.8518 (16). For share 0.3 it was 8.3518 (8).

Observation 5.3

what is not said. Finite-population corrections, combination with strata, costs of travel between plots, varying the number per plot, and how many trees a real survey should measure are not treated. What is shown is that, for the set distribution and costs, the best number does not depend on the budget and the loss from missing it follows a formula in one share.

6. Roll Call and Closing

1. variance formula for two-stage sampling (Section 1): between- and within-plot variance.
2. best **15.8518** trees (share 0.1, cost 20) (Computation 2.1): exhaustive 16 for all three budgets.
3. the same **1.123327** at twice and half (Computation 3.1): the loss is a formula in α and the factor.
4. **4.432163** with one tree (Computation 3.1): far from the best as a factor.

5. **0.983340 to 1.033172** of the formula by random sampling (Computation 4.1): as the formula says.

6. **between plots, or within** (Observation 5.1): the separator; the scope notes of Paper 957 and Paper 938.

Closing: The best number of trees to measure per plot ignores the budget, and doubling or halving it costs the same. Splitting the volume distribution of Paper 957 into plot and tree effects, with plot-effect share 0.1 and plot cost 20 the best was 15.8518 trees, and exhaustive search gave 16 for budgets of 1000, 2000 and 5000: twice and half the best gave the same 1.123327 times the variance, and the loss matched a formula in the share of the plot cost taken by the trees' total cost and the factor alone. Random sampling gave 0.983340 to 1.033172 times the formula. The separator is whether the variation lies between plots or within them. Placed among the earlier papers: to the forest samples Paper 957 and Paper 938 counted by stratum boundaries and allocation, the subsampling within plots and costs that neither treated were added. To be honest. This is one distribution with an infinite population, comparing a variance formula with random sampling. What sets the best number and how much is lost by missing it, and no further.

References

1. Cochran, W. G. (1977). *Sampling Techniques*. Wiley.
 2. Yamagishi, Y. (2026). *Stratifying a Forest Removes Between-Stratum Differences from Sampling Error, and Allocating More Plots to Variable Strata Reduces It Further* [Paper 938]. YOI: <https://www.yamagishiyuuki.com/yoi/938>
 3. Yamagishi, Y. (2026). *Stratum Boundaries at Equal Steps of Cumulative Root Frequency Come Close to Optimal, While Equal Counts Do Worst* [Paper 957]. YOI: <https://www.yamagishiyuuki.com/yoi/957>
-

— *From thinking to syncing.* —

Yuuki YAMAGISHI (山岸 勇輝)

© 2026 Yamagishi Complexity Torus Lab, Japan · Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0)