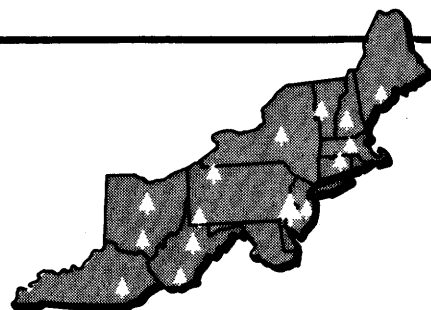


1973

Northeastern Forest Experiment Station



FOREST SERVICE, U.S. DEPT. OF AGRICULTURE, 6816 MARKET STREET, UPPER DARBY, PA. 19082

SAMPLING PLANTATIONS TO DETERMINE WHITE-PINE WEEVIL INJURY

Abstract.— Use of 1/10-acre square plots to obtain estimates of the proportion of never-weeviled trees necessary for evaluating and scheduling white-pine weevil control is described. The optimum number of trees to observe per plot is estimated from data obtained from sample plantations in the Northeast and a table is given of sample size required to achieve a standard error of the proportion of trees never weeviled of ± 0.05 , a value that can be obtained at a reasonable cost for the purpose intended.

Methods for evaluating and scheduling white-pine weevil control have been developed (*Marty and Mott 1964*) and field-tested (*Ford et al. 1965*). To use the evaluation scheme, an estimate of weevil intensity is needed. This paper is a report on the use of 1/10-acre plots to obtain estimates of the proportion of never-weeviled trees in plantations.

Methods and Results

Five 1/10-acre square plots were established in each of 32 white pine plantations from Maine to Virginia. The plots were widely distributed in each plantation so all weevil conditions would be represented in the sample. In each plot, every live tree was observed for evidence of weeviling, past or present; the total number of weeviled and never-weeviled trees was recorded for the plot.

Variance.—In sampling for proportions (p) with plots (cluster sampling), the mean and variance between plots do not have the simple binomial relationship expected from simple random sampling; but for a given size and

shape of plot there is sometimes a common multiplier or clustering factor (k) by which the between-plot variance can be approximated from the binomial variance.

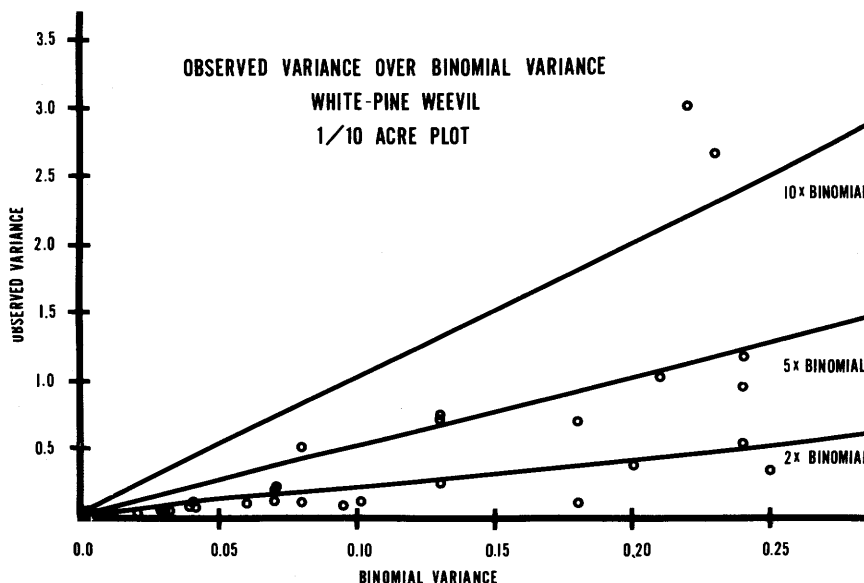
To investigate this possibility, mean squares between and within plots for each plantation were estimated by analysis of variance. The within-plot mean squares approximate the binomial variance ($\bar{p}[1-\bar{p}]$) very closely. The between-plot mean squares, like the binomial variances, are smallest for both small and large values of $\bar{p}$ and are largest in the mid-range of $\bar{p}$, but they exhibit an erratic pattern and are, in general, larger than the binomial variances.

When the between-plot mean squares are plotted over the binomial variances (fig. 1), it is clear that a common clustering factor will not represent all of the plantations. But, except for two aberrant values (which we cannot explain), the factors do have a relatively narrow range. In fact, 28 of the 32 plantation values have a factor of 5 or less; and for 21 plantations the factor was 2 or less. In the absence of other information about variances,

Figure 1.—Observed variance,

$$MS_b = M \sum_{i=1}^n (\bar{p}_i - \bar{\bar{p}})^2 / (n - 1),$$

of never-weeviled trees compared with binomial variance, $\bar{p}(1-\bar{p})$. The lines indicate various values of the clustering factor, k .



we suggest that the binomial variance and a clustering factor of 4 will usually be satisfactory for the design of cluster sampling plans using square 0.1-acre plots.

Cost relationships.—Efficient sample design also depends upon cost relationships among the component parts of the sampling process; in this case, upon the ratio of the cost (c_c) of locating and laying out a cluster (plot) to the cost (c_e) of locating and observing an element (tree) within the cluster.

Given the total man-hours for cluster sampling (C), the number of clusters in the sample (n), and the average number of elements observed per cluster (m) from each of about 10 plantations of varying densities (trees per acre), the component costs can be estimated by simple linear regression:

$$C/n = c_c + c_e(m)$$

The crude cost data at our disposal are not adequate for estimating c_c and c_e , but they do give evidence that the ratio c_c/c_e may have an average value between 50 and 100.

Application to Sampling Design

Sub-sampling fraction.—Though it is a common practice in cluster sampling to observe all the elements in a cluster, it is not necessarily an optimal practice. The optimum sub-sampling fraction (f_e) depends upon cluster size or number of elements per cluster (M), the clustering factor (k), and the cost ratio (c_c/c_e) in the following way:

$$f_e = m_{opt}/M = \frac{1}{M} \sqrt{\frac{M}{k-1}} \sqrt{\frac{c_c}{c_e}}$$

in which m_{opt} is the optimum number of elements to be observed in a cluster. This was modified from Sampford (1962, p. 176, equation 8.40), assuming that $S_w^2 = p(1-p)$ and $S_b^2 = kS_w^2/M$.

Optimum sub-sampling fractions were computed throughout the parameter space for the ranges of values we can reasonably expect will be encountered in practice: cluster size from 40 to 160 trees, clustering factor from 3 to 5, and cost ratio from 50 to 100. Sub-sampling

fractions between $\frac{1}{3}$ and $\frac{2}{3}$ encompassed almost the whole space, with fractions near $\frac{1}{2}$ occupying the central portion of the space. Because moderate departures from the optimum have only a small effect on sampling efficiency and because of its simplicity in practice, we suggest that the sub-sampling fraction of $\frac{1}{2}$ be used as standard. For each cluster, choose to examine either the first or the second tree encountered by a flip of a coin, and then examine every second tree thereafter in a systematic manner.

Cluster sampling fraction. — The fraction (f_c) of the total population of clusters (0.1-acre plots) to be included in the sample for an estimate of the proportion of trees never-weeviled to a predetermined level of precision may now be computed:

$$f_c = n/N = \frac{p(1.0-p) [k/M + (1.0-f_c)/m]}{p(1.0-p)k/M + N(SE)^2}$$

in which

- n = The number of clusters in the sample.
- N = The number of clusters in the population; in this case, the number of acres in the plantation multiplied by 10.
- p = The anticipated proportion of never-weeviled trees.
- $k = 4$ = The clustering factor assumed.
- $f_c = \frac{1}{2}$ = The near-optimal sub-sampling fraction.
- SE = The desired standard error of the estimated proportion of never-weeviled trees.

This was modified from Sampford (1962, p. 176, equation 8.39), assuming that $S_w^2 = p(1-p)$ and $S_b^2 = kS_w^2/M$.

We have tabulated (table 1) the sample sizes required to achieve a standard error of the proportion of trees never-weeviled of ± 0.05 , a value which can be obtained at reasonable cost and is suitable for the purpose intended.

Table 1.—The number of 0.1-acre plots required for estimating the proportion of trees never-weeviled at specified values for plantation of 1 to 5 acres and 5 to 20 acres
[SE = 0.05, $k = 4$, $f_c = \frac{1}{2}$]

M	P								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	10 < N < 50				0.1-ACRE PLOTS				
121	2	3	4	4	4	4	4	3	2
89	2	4	4	5	5	5	4	4	2
68	3	4	5	6	6	6	5	4	3
54	3	5	6	7	7	7	6	5	3
	50 < N < 200				0.1-ACRE PLOTS				
121	2	3	4	4	5	4	4	3	2
89	2	4	5	6	6	6	5	4	2
68	3	5	6	7	7	7	6	5	3
54	4	6	8	9	9	9	8	6	4

Literature Cited

- Ford, R. P., R. L. Talerico, and D. G. Mott.
1965. A FIELD TEST OF PROCEDURES FOR EVALUATING
AND SCHEDULING WHITE-PINE WEEVIL CONTROL.
USDA Forest Serv. Res. Note NE-37. 4 p.
- Marty, R., and D. G. Mott.
1964. EVALUATING AND SCHEDULING WHITE-PINE
WEEVIL CONTROL IN THE NORTHEAST. USDA Forest
Serv. Res. Pap. NE-19. 56 p.
- Sampford, M. R.
1962. AN INTRODUCTION TO SAMPLING THEORY WITH
APPLICATIONS TO AGRICULTURE. Oliver and Boyd,
London. 292 p.

—ROBERT L. TALERICO and ROBERT W. WILSON, JR.

Research Entomologist and Research Forester
Forest Service, U.S. Department of Agriculture
Northeastern Forest Experiment Station
Forest Insect and Disease Laboratory
Hamden, Connecticut

This study was a joint effort by the USDA Forest Service Northeastern Area State and Private Forestry office and the Northeastern Forest Experiment Station to improve the white-pine weevil evaluation guide. The authors acknowledge the assistance of Forest Pest Management, Northeastern Area State and Private Forestry, in the collection of field data.
