ESTIMATION OF STRATA AREAS IN FOREST SURVEY

AARNE NYYSSÖNEN and PEKKA KILKKI

HELSINKI 1966

Foreword

This paper is of those concerned with studies of forest survey methods at the Institute of Forest Mensuration and Management, University of Helsinki. The project has in part been financed with the aid of a grant made by the United States Department of Agriculture, Agricultural Research Service. The Society of Forestry in Finland has also given financial support to the project.

The authors have consulted each other at various stages of the work. *Nyyssönen*, the senior author, gave the theme with test materials to *Kilkki*, the junior author, who worked rather independently in the treatment of data, and writing a master's thesis on the subject. From this, the final paper was worked up and written by *Nyyssönen*.

Among the persons who helped in the work the authors wish to give especial mention of Mr *Erkki Mikkola*, M.Sc., for advice on statistical treatment.

Contents

1. Introduction

One important goal in making a forest inventory is classification of the division of a forest area into different strata. The information obtained in respect of these strata can then be of value also in more precise estimation of the total growing stock, for instance through stratified sampling. However, this stratification presupposes that the area of the strata can be estimated to an adequate degree of accuracy.

As a rule, the greatest accuracy in determining the stratum-areas is achieved when these are delineated on a map and measured with the planimeter. As this is a time-consuming task, it is frequently necessary to determine the areas on the basis of samples taken in the field, or from aerial photographs. Systematic sampling is commonly used for the purpose. Thus, the distribution of sample plots among strata can be employed; these plots are often placed at equal distances in both directions, as in uniform systematic plot sampling. Furthermore, a very common method applied in estimation of the stratum-areas has been that of measuring the lengths of parallel survey lines belonging in different strata. This transect method can naturally also be applied by measuring survey tracts instead of parallel lines (cf. HAGBERG 1957).

It should be remarked that no general equation is available for calculation of the standard error in systematic sampling. Assessment of the precision accordingly presents a problem. The methods developed for random sampling have been commonly applied to systematic plot sampling (cf. WILSON 1949; BICKFORD 1952; LOETSCH and HALLER 1964). Then again, some studies have been concerned with the precision of sampling with equidistant lines to determine the area. OSBORNE (1942) used the correlation between lines as one factor in his equation. The method is correct in theory, although in practice it calls for a great deal of work without an assurance that the estimate of correlation is accurately determinable. With equations presented by MATÉRN (1960), and based upon distance correlation, SEIP (1964) compiled a table which indicated the standard error in line sampling. He also prepared a table for plot sampling on the basis of binomial distribution.

It seems that the problem of determining the areas through systematic plot or line sampling is far from satisfactory solution. Since research work is consequently urgently needed, the aim of the present study is that of developing

methods suitable for calculation of the reliability of results obtained with regard to the stratum-areas by uniform systematic plot sampling and sampling with equidistant lines (strip survey). Assessment of the factors which influence the standard error of the stratum-areas obtained by the sampling methods mentioned has been attempted here mainly by means of regression analysis. Throughout the whole study, it has been assumed that the total area is known.

2. Test materials

The investigation was carried out in two experimental areas, one located in southern, and the other in northern Finland. In addition, a third area in the central part of the country was available to test the results arrived at. In all of these areas, stands were delineated on a map, and information on the distribution of strata was available.

Of the two study areas proper, the smaller one is located in the E v o forestry district (61° 15' N.lat., 25° 10' E.long.), and is 100 hectares in size. Fig. 1 is a stand map of the area. The following list gives the percentages of different strata, and shows the strata to which the different stand figures belong. So-called treatment classes were used as strata; the non-forest areas were combined in one stratum.

Stratum 1:   2 per cent, No.   12, 63, 95
     »   2: 12 » » »   08, 16, 21, 27, 43, 49,
                                        56, 58, 64, 68, 96
     »   3:   7 » » »   10, 22, 23, 24, 25, 31,
                                        44, 50, 57, 75, 93, 98
     »   4: 24 » » »   01, 03, 05, 06, 09, 14,
                                        20, 35, 39, 41, 42, 45,
                                        46, 48, 52, 54, 55, 65,
                                        70, 71, 72, 76, 80, 82,
                                        83, 84, 85
     »   5: 40 » » »   04, 07, 11, 15, 26, 28,
                                        29, 33, 36, 37, 38, 47,
                                        53, 60, 62, 66, 67, 69,
                                        73, 74, 78, 81, 88, 89,
                                        94, 97, 99
     »   6:   2 » » »   40, 87, 90
     »   7:   1 » » »   17, 51
     »   8:   2 » » »   13, 19, 32
     »   9: 10 » » »   02, 18, 77, 79
     » 10:   0 » » »   91, 92

The area comprises 10 000 squares, each measuring 100 square metres; the stands have been formed by combining such squares.
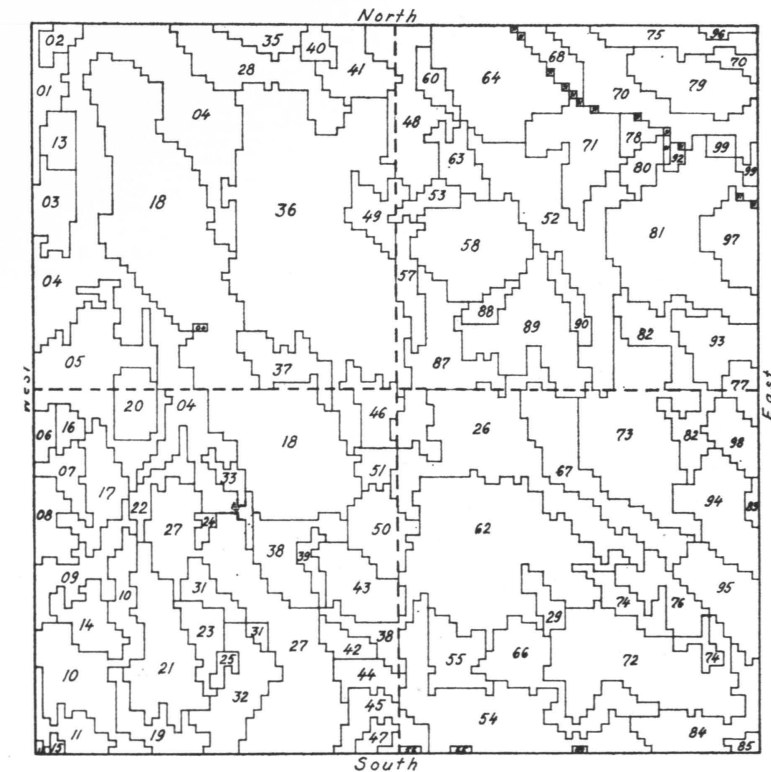
Fig. 1. Evo area. Scale about 1: 10 000.

The area was sampled by both uniform systematic plot sampling and sampling with equidistant lines, using samples of varying size.

The total number of 10 000 squares formed the population in the *plot sampling*. When a certain plot distance was used, the number of samples obtained was K; this can be solved by the equation $K = N/n$, in which $N =$ the size of the population, and $n =$ the size of the sample. In order to make K an integer, N was often less than 10 000.

In each sample there was counted the number of plots in each stratum. On the basis of sample means relating to a certain type of sample, the standard error of the mean of plot number (s_{pn}) in each stratum was calculated by the equation

$$s_{pn} = \sqrt{\frac{\sum\limits_{i=1}^{K}(p_i n)^2 - \dfrac{(\sum\limits_{i=1}^{K} p_i n)^2}{K}}{K}} \qquad (1)$$
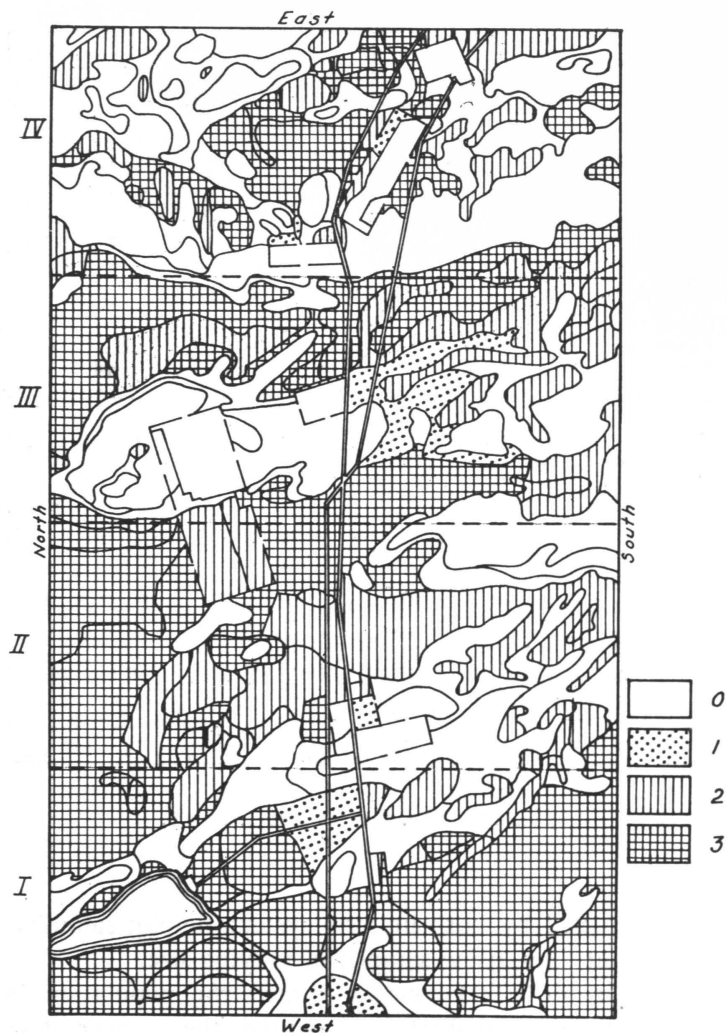
Fig. 2. Meltaus area. Scale 1: 30 000.

Group 0 = Strata 7, 8 and 9 combined
»   1 = Stratum 1
»   2 = Strata 2 and 6 combined
»   3 = Strata 3, 4 and 5 combined.

where p_i = relative size of stratum p in sample i, and n = size of the sample.
The result represents one »observation» employed in making a regression analysis.

K was used instead of K-1, since the samples together formed the population.
The number of plots in one sample ranged from 16 to 625.

The population of *survey lines* comprised 100 strips 10 metres wide. The

Table 1. Strata percentages in the Meltaus area.

Sub-area	Stratum									Total
	1	2	3	4	5	6	7	8	9	
	Percentage of the area									
I	5	1	37	18	5	5	13	9	7	100
II		17	13	22	—	17	16	12	3	100
III	7	12	24	7	—	10	15	19	6	100
IV	1	2	4	16	2	13	34	23	5	100

number of samples in the line sampling was solved in the same manner as in the plot sampling. The number of survey lines per sample ranged from 2 to 25. K was an integer also in this case. By replacing n in Equation (1) by the line length in the sample, the result indicates the standard error of the mean length of the lines in stratum p. The same procedure was applied in the following area as well.

The larger area (about 900 hectares) is called M e l t a u s (67° N.lat., 25° 20′ E.long.). The area is rectangular in form, with the longer side 1.7 times the shorter one. The natural orientation of the long axis of the stands in the area is clearly from north to south, as is discernible in Fig. 2, where nine primary strata have been combined in four groups.

The original stand map of the Meltaus area available had a scale of 1: 15 000. On this map, the area was divided by north to south lines into four sub-areas of equal size. Table 1 shows the division of these sub-areas into the strata. In each sub-area, lines at intervals of one millimetre, thus representing a 15-metre wide strip, were located in both directions. In this way, two line systems covering the area were obtained. The distribution of the lines on different strata was measured to an accuracy of one millimetre. In all, 264 N-S lines and 153 E-W lines were measured.

In treatment of the material, eight different areas were formed from the four sub-areas as follows:

Area 1: Sub-area   I
 »   2:   »   II
 »   3:   »   III
 »   4:   »   IV
 »   5: Sub-areas   I and II
 »   6:   »   III and IV
 »   7:   »   I, II, and III
 »   8:   »   I, II, III, and IV.

As in Evo, systematic samples were taken on different areas from the line systems measured. In order to make K an integer, the areas with the N-S lines

were so reduced that the number of lines crossing them was divisible by 64; 150 of the E-W lines were used in the study. The number of lines in one sample ranged between 2 and 50. Strata 1 and 2, along with 3 and 4 on the N-S lines were combined, in view of the limited capacity of the electronic computer.

On both of the test areas, the data relevant to the distribution of the plots and lines in different strata were punched on cards. A program was worked out for the computer, which was capable of selecting and handling the data needed in surveys with different line and plot distances. The calculations were mainly effected on the IBM 1620 computer at the Computing Centre of the Department of Mathematics, University of Helsinki.

The results of the first stage of calculation consisted of data on the standard error and other characteristics of the samples. The examination of these data was based upon regression analysis.

In analyses of the data from plot sampling, there were discarded those cases in which the number of plots or stratum areas was so small that it completely failed to appear in at least 10 samples. By this means, an attempt was made to avoid skewness in the distribution of the p_t-values which might possibly be caused by 0-samples. The final number of observations was 81. In line sampling, again, all the cases including 0-samples were rejected. Thus, for regression analyses, there were 90 observations available from Evo; 43 on N-S lines and 47 on E-W lines. In Meltaus, each of the eight areas formed provided 80 to 98 observations (23 to 34 on N-S lines and 53 to 70 on E-W lines), or in total 730 observations (217 on N-S and 513 on E-W lines).

3. Calculation of regression equations

31. Uniform systematic plot sampling

If it is assumed that the area of a given stratum is to be determined by random sampling, i.e., with plots located randomly in the field, on maps, or on aerial photographs, the standard error of the number of plots falling on the stratum (s_{pn}) can be calculated by the equation (cf. Cochran 1963):

$$s_{pn} = \sqrt{\frac{(N-n)\,pqn^2}{(n-1)\,N}} \qquad (2)$$

where N = size of the population,
       n = size of the sample,
       p = relative size of the stratum,
       $q = 1 - p$

When the size of sample increases, $n-1$ can be replaced by n without reducing the reliability of the result. The equation takes the form:

$$s_{pn} = \sqrt{\frac{(N-n)\,pqn}{N}} \qquad (3)$$

When, in addition, the size of the population increases, $\dfrac{N-n}{N}$ approaches one, and there will be the form

$$s_{pn} = \sqrt{pqn}$$

The standard error of the number of plots in a stratum can also be estimated by means of the above equations in the case when uniform systematic plot sampling is employed, if it is assumed that the stratum is randomly distributed in an area. Although this assumption as such is hardly ever valid, it may be presumed that the standard error in uniform systematic plot sampling depends on the same variables as in random sampling. Moreover, it can be assumed that the dependence of the standard error on these variables is approximately of the same type in both cases.

Equation (3) can also be written in the following form:

$$s_{pn} = 1\,(pq)^{1/2}\,n^{1/2}\,(1-n/N)^{1/2} \qquad (4)$$

When the constants in the equations are replaced by terms a, b, c, and d, and the logarithms are taken on both sides, there is obtained the following function model applicable in the regression analyses:

$$\log s_{pn} = \log a + b\log (pq) + c\log n + d\log (1-n/N) \qquad (5)$$

In the regression analyses carried out on the basis of the Evo data of the precision of uniform systematic plot sampling, use was made solely of the variables occurring in Equation (5).

Table 2 presents the results of the regression analyses made. The correlation

Table 2. Dependent variable $\log s_{pn}$ as explained by different combinations of independent variables in plot sampling in Evo.

Independent variable			R	S as percentage
$\log (pq)$	$\log n$	$\log (1-n/N)$		
—	—	—	0	83.4
×	—	—	0.796	50.8
×	×	—	0.904	36.1
×	×	×	0.909	35.4

coefficient (R) and the standard error of estimate (S) provide information on the mutual relationships of the different equations, and not directly on the final precision of the results. Since $\log n$ and $\log (1 - n/N)$ are in a relatively strong negative correlation $(r = -0.893)$, the addition of $\log (1 - n/N)$ as an independent variable improves the result no more than slightly. The last equation takes the following form:

$$\log s_{pn} = 0.0433 + 0.419 \log (pq) + 0.348 \log n + 5.23 \log (1 - n/N) \qquad (6)$$

For regression coefficients, the t-values are 20.1, 6.3, and 2.0.

32. Sampling with equidistant lines

As it can be assumed that the survey lines are formed by sample plots placed one after another, it seems possible also to characterize the precision of sampling with equidistant lines, on the basis of the same function models as in uniform systematic plot sampling. Then s_{pn} expresses the standard error of the total length of lines in stratum p, and n the total length of lines in the sample. The additional variables used here were $\log N$, and some characteristics meant to describe the distribution of the strata in the area. First, the characteristic s_{pl} refers to the standard deviation of the length of lines in stratum p (cf. Fig. 3). Since the standard error of the line length in random sampling is directly proportional to s_{pl}, it can be assumed that the accuracy of stratum-
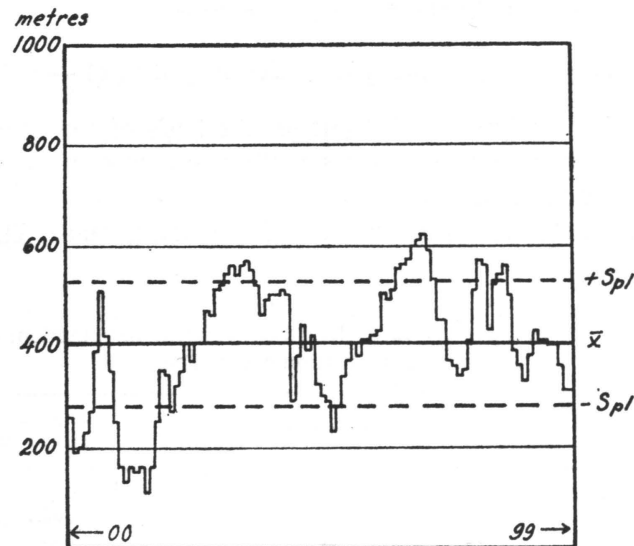

Fig. 3. Amount of treatment class 4 on 100 N-S strips in Evo, with mean ($\bar{x}$) and standard deviation (s_{pl}).

areas obtained by sampling with equidistant lines can be significantly dependent on s_{pl}. Secondly, the average number of lines crossing the strata, and the effective number of lines, as recommended by KUUSELA (1960), may be applied. The symbol m_p will be used for the former, and the latter has here been calculated by the equation

$$l = \frac{pn}{\dfrac{pn}{m_p} + s_{pl}}$$

In other sampling studies carried out simultaneously with the present investigation, the hypothesis that the variance of a systematic sample provides in general a rather good estimate of the population variance, has been proved true. This finding makes it possible to use s_{pl}, calculated for the population, as a variance for individual samples in the regression analyses.

Regression analyses of the Evo data were made for both the lines of different directions separately, and for the total material. In this connection, attempts were made to test different combinations of independent variables, although $\log N$ was excluded.

The principal results of the regression analyses of the Evo material are presented in Table 3. The characteristic of the relative size of stratum, $\log (pq)$, bears a marked correlation with the dependent variable. The characteristics of the distribution of stratum-areas, $\log (mp)$, $\log l$ and $\log s_{pl}$, improve the results to some extent, but good results, from a relative standpoint, were derived by means of the function in the last row:

$$\log s_{pn} = 1.483 + 0.441 \log (pq) + 2.17 \log (1 - n/N) \qquad (7)$$

The t-values of the regression coefficients are 12.4 and 4.9.

The regression equations from the Meltaus material were initially calculated

Table 3. Dependent variable $\log s_{pn}$ as explained by different combinations of independent variables in strip sampling in Evo.

Independent variable						R	S as percentage
$\log (pq)$	$\log n$	$\log (1 - n/N)$	$\log m_p$	$\log l$	$\log S_{pl}$		
—	—	—	—	—	—	0	42.5
×	—	—	—	—	—	0.816	24.7
×	×	×	×	—	—	0.861	22.1
×	×	×	—	×	—	0.862	22.0
×	×	×	—	—	×	0.864	21.9
×	—	×	—	—	—	0.860	21.9

Table 4. Dependent variable log s_{pn} as explained by different combinations of independent variables in strip sampling in Meltaus.

log (pq)	log n	$\frac{\log}{(1-n/N)}$	log m_p	log l	log S_{pl}	log N	R	S as percentage
—	—	—	—	—	—	—	0	29.9
×	×	—	—	—	—	×	0.719	20.8
×	×	×	—	—	—	×	0.720	20.8
×	×	×	×	—	—	×	0.733	20.4
×	×	×	—	×	—	×	0.746	20.0
×	×	×	—	—	×	×	0.759	19.5
×	×	—	—	—	×	×	0.759	19.5

separately for different line directions and for each of the eight areas, but it seemed evident that the whole material could be handled as one unit without losing much of the explanatory ability of the equations. This offered the considerable advantage of increased range of variation.

Table 4 indicates the results of the regression analyses made for the whole material. In choosing the equations for this table, attention was paid to the predicting strength of different variables.

Among the equations calculated, that in the last row of Table 4 has the following form:

$$\log s_{pn} = -0.479 + 0.231 \log (pq) - 0.617 \log n + 0.518 \log s_{pl} + 0.412 \log N \tag{8}$$

The coefficients of the independent variables used in Equation (8) differ from zero to a highly significant degree, as their t-values range from 6.6 to 10.0.

Log m_p and log l also proved to be significant variables when log s_{pl} was not included. It appears, however, that log s_{pl} alone is capable of compensating for both of them.

As the Meltaus material includes eight different areas, within each of which the sampling was done in two directions, the range of variation of the characteristics describing the distribution of strata is considerably larger than that of the Evo material. As a consequence, the significance of these characteristics as independent variables is brought out more clearly than in the Evo material, which comprises one area alone. In the regression analyses calculated for the Meltaus area, one important independent variable is the size of the area, which could not be analysed at all on the basis of the Evo material.

4. Estimation of the precision of stratum-area determination

41. Sampling with random plots

In plot sampling, the standard error of a stratum-area (s_{sa}) is derived by multiplying the standard error of the number of plots falling in that stratum by A/n, where A is the total area. Thus, if the plots are distributed randomly, Equation (9) is derived from Equation (2), p. 10:

$$s_{sa} = \sqrt{\frac{(N-n)\,pq}{(n-1)\,N}}\,A \tag{9}$$

If the sampling unit is conceived as a point, with the consequence that an infinite number of them can be placed in the area being estimated, the following equation is applicable:

$$s_{sa} = \sqrt{\frac{pq}{n}}\,A$$

42. Uniform systematic plot sampling

The degree of precision of estimation of an area by uniform systematic plot sampling is obtained, for an area 100 hectares in size, by multiplying the standard error of the plot number of a certain stratum (Equation 6, p. 12) by $\frac{100}{n}$. When furthermore the addition required by the logarithmic transformation (cf. JEFFERS 1960) is made, the following equation is obtained:

$$s_{sa} = 1.13\,(pq)^{0.419}\,n^{0.348}\left(1 - \frac{n}{10000}\right)^{5.23}\frac{100}{n}$$

In this equation, n can be replaced by $\frac{1000000}{v^2}$, in which v indicates the plot distance in metres. The equation takes the shape:

$$s_{sa} = 0.0139\,(pq)^{0.419}\,v^{1.30}\left(1 - \frac{100}{v^2}\right)^{5.23} \tag{10}$$

By way of reservation, it can be assumed that the precision of plot sampling changes in the same proportion as that of line sampling with changing total area. Some equations calculated from the Meltaus data allow of derivation of the effect of the change in area on the precision of line sampling. These equations give the result that when the area changes k-fold, the standard error of the stratum-area changes $k^{0.488}$-fold, if the line distance remains unchanged.

Table 5. Standard error of a stratum-area in uniform systematic plot sampling, based on Equation (11).

Plot distance, metres	Relative size of stratum, per cent									
	10 / 90	20 / 80	30 / 70	40 / 60	50	10 / 90	20 / 80	30 / 70	40 / 60	50
	Standard error, hectares									
	Total area 100 hectares					Total area 200 hectares				
50	.7	.8	.9	1.0	1.0	.9	1.2	1.3	1.4	1.4
100	1.9	2.4	2.8	3.0	3.0	2.8	3.5	3.9	4.1	4.2
150	3.4	4.3	4.8	5.1	5.2	4.8	6.0	6.8	7.1	7.3
200	5.0	6.3	7.1	7.5	7.6	6.9	8.9	10.0	10.5	10.7
250	6.7	8.6	9.6	10.1	10.3	9.4	11.9	13.4	14.2	14.4
	Total area 500 hectares					Total area 1000 hectares				
100	4.3	5.4	6.0	6.4	6.5	6.0	7.6	8.5	9.0	9.2
200	10.9	14.0	15.6	16.5	16.7	15.3	19.6	21.8	23.2	23.6
300	18.7	23.8	26.6	28.2	28.7	26.2	33.4	37.4	39.6	40.2
400	27.2	34.7	38.9	41.1	41.8	38.2	48.7	54.6	57.6	58.6
500	36.5	46.4	52.0	55.1	56.0	51.2	65.2	73.0	77.2	78.5

When the size of the area is A hectares, the standard error of the stratum-area is obtained by multiplying Equation (10) by $(\frac{A}{100})^{0.488}$. When the equation obtained in this way is further simplified, it takes the form:

$$s_{sa} = 0.00146 \, (pq)^{0.419} \, v^{1.30} \left(1 - \frac{100}{v^2}\right)^{5.23} A^{0.488} \tag{11}$$

Table 5 presents figures calculated on the basis of Equation (11).

43. Sampling with equidistant lines

For areas 100 hectares in size, the standard error of a stratum can be estimated in line sampling on the basis of Equation (7) p. 13, derived from the Evo material. This equation is based upon a more exact map than is, for instance, the corresponding equation of the Meltaus material. Only the relative size of the stratum of the total area needs to be known, along with the total length of the survey line. In Equation (7), the length of the sampling line is replaced by $\frac{100000}{v}$ where v = the line distance expressed in metres; moreover, Equation (7) is converted into a form which gives the error in hectares.

Table 6. Standard error of a stratum-area in strip sampling, based on Equation (13).

Line distance, metres	Relative size of stratum, per cent									
	10 / 90	20 / 80	30 / 70	40 / 60	50	10 / 90	20 / 80	30 / 70	40 / 60	50
	Standard error, hectares									
	Total area 100 hectares					Total area 200 hectares				
50	.3	.4	.5	.5	.6	.5	.6	.8	.8	.8
100	.9	1.2	1.3	1.4	1.4	1.3	1.6	1.8	1.9	1.9
150	1.5	1.9	2.2	2.3	2.3	2.1	2.7	3.0	3.1	3.2
200	2.1	2.6	2.9	3.1	3.2	2.8	3.7	4.1	4.4	4.4
250	2.6	3.4	3.8	4.0	4.1	3.7	4.8	5.3	5.6	5.7
	Total area 500 hectares					Total area 1000 hectares				
100	1.9	2.6	2.9	3.0	3.1	2.8	3.6	4.0	4.3	4.3
200	4.4	5.7	6.5	6.9	7.0	6.3	8.1	9.1	9.6	9.8
300	6.9	9.0	10.1	10.7	10.9	9.7	12.5	14.2	15.0	15.4
400	9.4	12.2	13.6	14.6	14.8	13.3	17.1	19.1	20.4	20.8
500	11.9	15.4	17.3	18.4	18.7	16.8	21.6	24.3	25.8	26.3

When the addition required by the logarithmic transformation is made, the following equation is obtained:

$$s_{sa} = 0.0329 \, (pq)^{0.441} \left(1 - \frac{10}{v}\right)^{2.17} v \tag{12}$$

If estimation of the standard error of the stratum-area is desired in areas other than 100 hectares in size, the result obtained from Equation (12) has to be multiplied by a correction coefficient in the same manner as in uniform systematic plot sampling. This gives the following equation:

$$s_{sa} = 0.00347 \, (pq)^{0.441} \left(1 - \frac{10}{v}\right)^{2.17} v \, A^{0.488} \tag{13}$$

Table 6 presents the results of calculation by this equation.

According to the above equations, the standard error of a stratum-area can be estimated in forest areas approximating the Evo area in form and pattern. Even in cases when the area differs considerably from that at Evo, these equations can probably give a concept of the effect of the relative size of a stratum, the line distance, and the size of the total area on the standard error of the stratum-area.

If, moreover, it is wished to use the characteristics of the individual pattern and form of each area in a study of the precision of line sampling, the equations

Table 7a. Standard error of a stratum-area in strip sampling of 100 hectares, based on Equation (15).

Line distance, metres	Relative size of stratum, per cent									
	10 / 90	20 / 80	30 / 70	40 / 60	50	10 / 90	20 / 80	30 / 70	40 / 60	50
	Standard error, hectares									
	$s_{pl_1} = 100$					$s_{pl_1} = 200$				
50	.4	.4	.5	.5	.5	.6	.7	.7	.8	.8
100	.9	1.0	1.1	1.1	1.1	1.3	1.4	1.5	1.6	1.7
150	1.5	1.7	1.8	1.8	1.9	2.1	2.4	2.5	2.7	2.7
200	2.0	2.3	2.5	2.6	2.6	2.9	3.3	3.5	3.6	3.8
250	2.6	3.0	3.2	3.3	3.3	3.8	4.3	4.6	4.8	4.8

calculated for the whole Meltaus material need to be employed; the best of them is probably No. (8), p. 14. With a view to making this equation simpler in application, the characteristics can be so converted that they correspond to measurements in the field instead of map measurements. As the scale of the map was 1: 15 000, the dimensions measured in metres in the field have to be multiplied by 0.06667 to provide the variables needed for the equation. The total length of the line can be replaced by the line distance. This gives the following equations:

$$n = \frac{666.7\ A}{v}$$

$$s_{pl} = 0.06667\ s_{pl_1}$$

$$N = 44.44\ A$$

in which s_{pl_1} indicates the distribution of the length (in metres) of stratum p on the lines. When the new variables thus obtained are inserted in Equation (8), the antilogarithms are taken, and the correction required by the logarithmic transformation is introduced, the following equation is obtained:

$$s_{pn} = 0.146\ (pq)^{0.231}\ v^{0.167}\ s_{pl_1}^{0.518}\ A^{0.245} \qquad (14)$$

This equation gives the standard error of the portion of the sampling line which falls on stratum p. The standard error of the corresponding stratum-area can be calculated by the equation:

$$s_{sa} = \frac{s_{pn}}{n}\ A = 0.0015\ v\ s_{pn}$$

Table 7b. Standard error of a stratum in strip sampling of 1 000 hectares, based on Equation (15).

Line distance, metres	Relative size of stratum, per cent									
	10 / 90	20 / 80	30 / 70	40 / 60	50	10 / 90	20 / 80	30 / 70	40 / 60	50
	Standard error, hectares									
	$s_{pl_1} = 100$					$s_{pl_1} = 200$				
100	1.5	1.9	2.0	2.0	2.0	2.3	2.7	2.8	2.9	2.9
200	3.5	4.1	4.3	4.4	4.5	5.1	5.9	6.2	6.4	6.5
300	5.8	6.6	7.0	7.2	7.3	8.3	9.4	10.1	10.4	10.5
400	8.1	9.2	9.8	10.2	10.3	11.6	13.2	14.0	14.6	14.7
500	10.5	11.9	12.7	13.2	13.3	15.0	17.1	18.2	18.8	19.0
	$s_{pl_1} = 300$					$s_{pl_1} = 400$				
100	2.9	3.2	3.4	3.5	3.5	3.3	3.8	4.0	4.1	4.2
200	6.3	7.2	7.7	8.0	8.1	7.3	8.3	8.8	9.2	9.3
300	10.3	11.6	12.5	12.8	12.9	11.8	13.5	14.4	14.8	14.9
400	14.4	16.4	17.4	18.0	18.1	16.6	19.0	20.2	20.8	21.0
500	18.5	21.1	22.6	23.2	23.4	21.5	24.6	26.1	27.0	27.2

When s_{pn} in this formula is replaced by its value in Equation (14), the following equation is obtained:

$$s_{sa} = 0.000219\ (pq)^{0.231}\ v^{1.17}\ s_{pl_1}^{0.518}\ A^{0.245} \qquad (15)$$

Tables 7a and 7b present the results of calculations by the application of this equation. The exponent of A in this equation differs from that in Equation (13); this is attributable to a part of the effect of the total area being brought forth by the exponent of s_{pl_1}.

5. Reliability of the results

The material obtained from the Evo area is well adapted for a comparison of line sampling and plot sampling. A weakness of this material is the relatively small size of the area (100 hectares). Nevertheless, the Meltaus material offers a considerably larger range of variation. In line sampling, the regression equations calculated for the whole Meltaus material are based upon 730 observations. The size and the form of the sub-areas measured there varied appreciably.

Some concept of the reliability of the regression models used is procurable by comparison of the results of different materials. Table 8 shows an example:

Table 8. Comparison of standard errors of a stratum-area in strip sampling, based on the Evo and Meltaus materials. Total area 100 hectares; $s_{pl_1} = 100$.

Line distance, metres	Relative size of stratum, per cent				
	10 / 90	20 / 80	30 / 70	40 / 60	50
	Standard error, hectares				
	Evo				
50	.37	.44	.48	.50	.50
100	.95	1.13	1.23	1.29	1.31
150	1.56	1.86	2.01	2.10	2.13
200	2.15	2.57	2.80	2.92	2.95
250	2.76	3.28	3.57	3.72	3.77
	Meltaus				
50	.40	.45	.49	.51	.51
100	.91	1.04	1.11	1.14	1.15
150	1.46	1.69	1.78	1.84	1.87
200	2.04	2.33	2.48	2.55	2.59
250	2.61	2.99	3.16	3.27	3.32
	Difference as percentage of the Meltaus results				
50	—8	—2	—2	—2	—2
100	4	9	11	13	14
150	7	10	13	14	14
200	5	10	13	14	14
250	6	10	13	14	14

the standard error of stratum-areas at varying line distance, calculated on the basis of Equation (8) of the Meltaus material, and of the corresponding equation of the Evo material. The size of the area is assumed to be 100 hectares, and the characteristic s_{pl_1}, for distribution of the stand figures, 100 metres. When the large differences between these two areas are borne in mind, the differences between the results arrived at from the two equations are surprisingly small in this example. Even the largest differences do not exceed 15 per cent.

Another means of comparing the reliability of the results achieved in line sampling is given by the stand map in Fig. 4. This was available on a scale of 1: 10 000 as regards the area 326.₇ hectares in size located in the Viitasaari forestry district, central Finland. Survey lines were measured at 1-millimetre distances in both NE-SW and NW-SE directions; the stands were classified in four strata, and their lengths on lines were measured to an accuracy of 1-millimetre (10 metres on the ground).
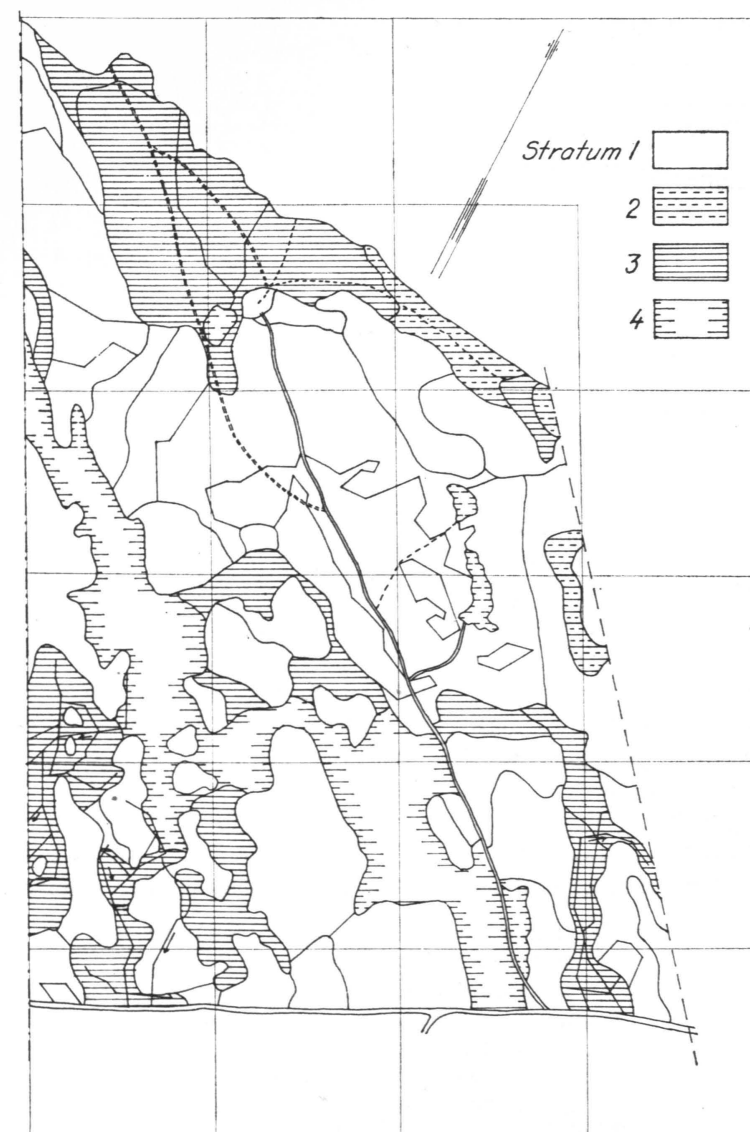
Fig. 4. Viitasaari area. Scale 1: 20 000.

For different line distances, three results were calculated in respect of each stratum:

1. the true standard deviations of the sample means;

2. the results given by Equation (13) p. 17, the independent variables being the relative size of stratum (p), the line distance (v) and the size of the total area (A);
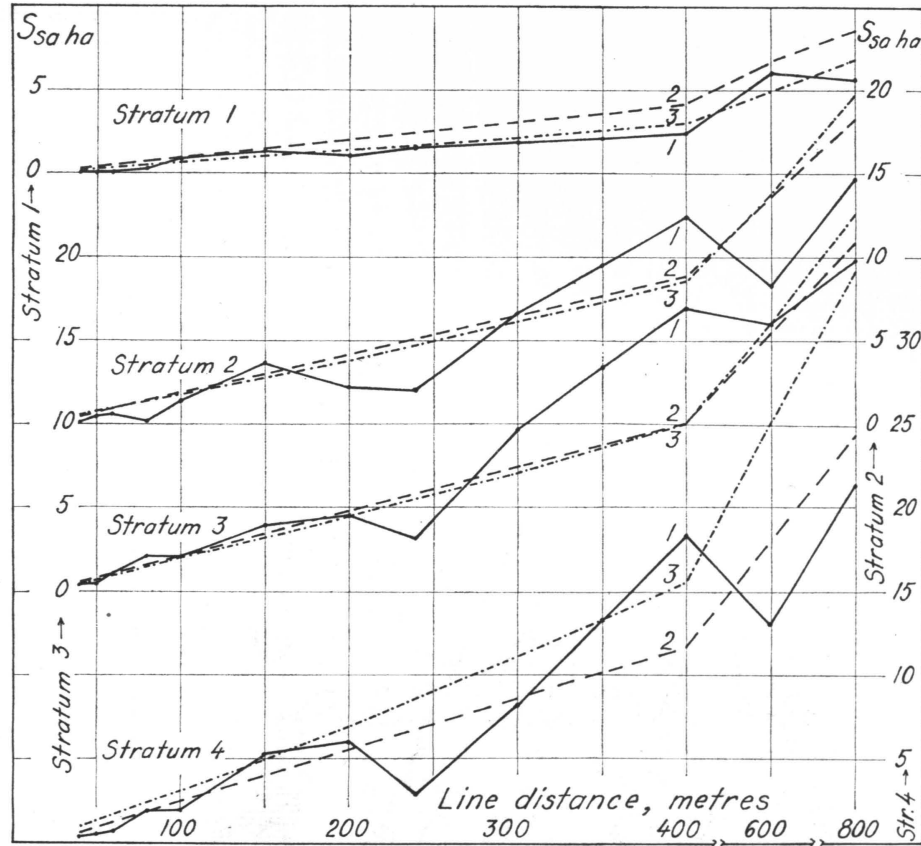
Fig. 5. Standard errors of stratum-areas from NE-SW lines in the Viitasaari area.

Curve No. 1: true standard deviations of sample means
»   »   2: standard errors estimated by Equation (13)
»   »   3: standard errors estimated by Equation (15).

Stratum 1	8.₀ hectares,	2.₅	per cent
» 2	47.₅ »	14.₅	»
» 3	70.₈ »	21.₇	»
» 4	200.₄ »	61.₃	»
Total area	326.₇ »	100	»

3. the results given by Equation (15) p. 19, with the standard deviation of the length of lines (s_{pl}) as an independent variable in addition to those mentioned above.

Fig. 5 indicates the results obtained from NE-SW lines, giving the regression of the standard error of the stratum-area on line distance. There are some features typical of systematic sampling in the course of the »true» standard errors, but as a whole the estimates of standard error given by Equations (13)

and (15) seem rather satisfactory; furthermore they are about the same at least as far as Strata 2 and 3 are concerned. Possibly, Equation (13) has given the best results for Stratum 4, and Equation (15) those for Stratum 1.

6. Discussion

The results derived from the Evo material permit of comparison being made between different sampling methods. Fig. 6 provides an example of such a comparison. For estimation of the standard error of stratum-areas, Equation (9) p. 15 was applied for random plot sampling, Equation (10) p. 15 for uniform systematic plot sampling, and Equation (12) p. 17 for sampling with equidistant lines. There are two scales on the abscissa of the figure. When systematic plot and line sampling are compared, the »distance» means plot or line distance. In comparing sampling with random or systematic plots, use can be made of the lower scale showing the number of plots.

Systematic plot sampling is superior to sampling with random plots. The relative merits of the former diminish as the number of plots is reduced. Calculation of the standard error of a stratum-area in systematic plot sampling, by application of the methods for random sampling, entails substantial overestimation of the error.

With an increase in the line distance, the advantages of line sampling become more obvious than those of uniform systematic plot sampling. This means that
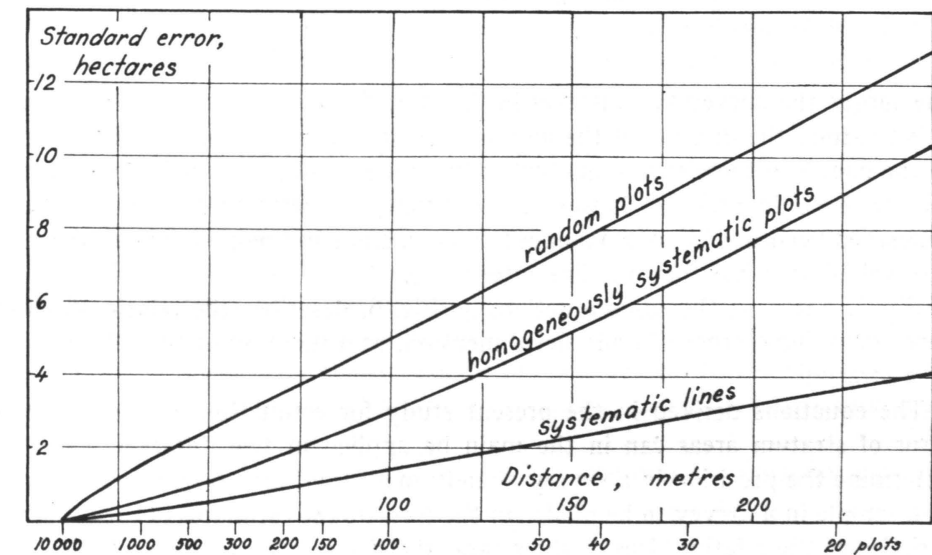

Fig. 6. Standard errors of a stratum on application of different sampling methods. Total area 100 ha., stratum-area one half of this.
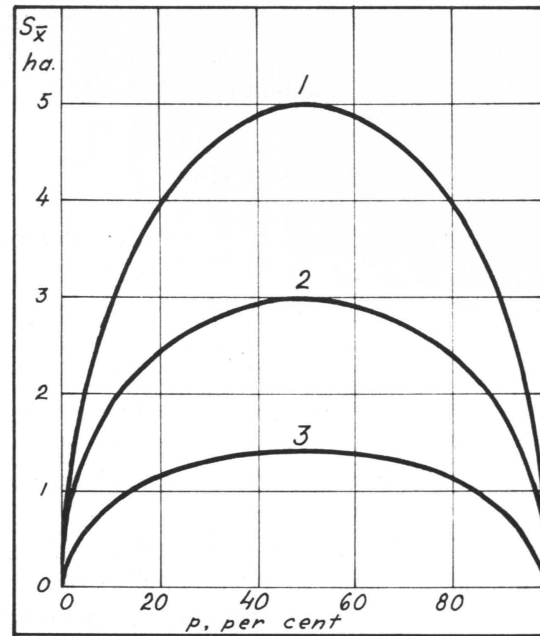
Fig. 7. Standard errors of stratum-areas on application of different sampling methods. Total area 100 ha.

Curve No. 1: sampling with 100 random plots
  »   »   2: uniform systematic plot sampling (100 plots)
  »   »   3: sampling with 10 equidistant lines

the larger the survey line interval in line-plot survey, the more advantageous it is to study the division of the area into strata by measuring the distribution of the lines into these strata, and not only by means of the sample plots located on the lines. In practice, however, the differences between the two methods concerned need not be so pronounced as here, since in line-plot survey the plot interval is often less than the line interval.

Fig. 7, based on the same functions as Fig. 6, describes the relationships of three sampling methods in another dimension, as a function of the relative size of a stratum.

The equations derived in the present study for estimation of the standard error of stratum areas can in the main be applied in two different ways: to determine the precision in a survey already made, and to predetermine the size of a sample in a survey to be made. In the former, the estimates of such characteristics as the relative sizes of the strata, the size of the sample, and even the distribution of the strata, are available in addition to the total area. The deviation of these estimates from their correct values reduces the reliability of the
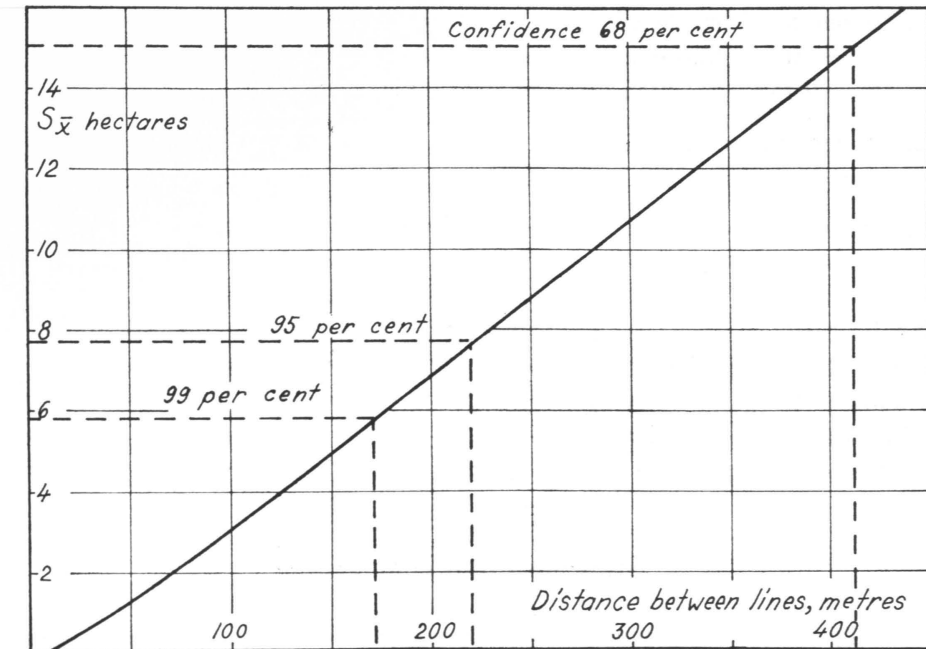
Fig. 8. Estimation of a line distance for a given precision: an example. Total area 500 ha., relative size of stratum 60 per cent, and allowable error 5 per cent, or 15 ha.

estimation of the standard error of the stratum-area. The greatest discrepancies in estimation of the standard errors occur when the size of the stratum is close to 0 or 100 per cent; under these conditions even a slight deviation from the correct size of the stratum-area induces a considerable deviation in the estimation of the error (cf. Fig. 7).

If one wishes to ascertain the size of the sample which will give the area of a stratum of stated size to a given degree of precision, those equations which do not call for information on the distribution of the stratum are often useful. Different sampling methods may be considered. Sampling with random plots has been described by LOETSCH and HALLER (1964, pp. 329—330). In uniform systematic plot sampling, and sampling with equidistant lines, the equations presented in the present paper may be applied. However, the use of Tables 5 and 6 may be found easier. For greater precision, graphic interpolation can be introduced.

An example of the estimation of the line distance required for a definite degree of precision is presented in Fig. 8, drawn from Table 6. The total area is 500 ha., the relative size of stratum 60 per cent, and the allowable error (AE) 5 per cent, equivalent to 15 ha. The standard error of the stratum was estimated for three confidence levels:

1. 68 per cent confidence

   $s_{sa} = AE = 15.0$ ha.; $v = 410$ m.

   ($v =$ distance between lines, from Fig. 8)

2. 95 per cent confidence

   $s_{sa} = AE/1.96 = 7.7$ ha.; $v = 220$ m.

3. 99 per cent confidence

   $s_{sa} = AE/2.58 = 5.8$ ha.; $v = 170$ m.

The methods of estimating confidence limits in the example give results which are almost correct when the p_i-values obtained from different materials are normally distributed. Both theoretically and experimentally, it could be concluded that the distribution of p_i-values is rather close to normal distribution, except when p approaches 0 or 1, or when a sample is very small.

7. Summary and conclusions

The highest degree of precision in determining the areas of different strata in forest survey is achieved when these are measured from a map by means of a planimeter. However, in practice the stratum-areas usually need to be determined on the basis of samples taken in the field, or from aerial photographs. The goal of the present investigation has been determination of the precision in stratum-area estimation on the application of different sampling methods.

The objects of study comprised three sampling methods: 1. sampling with random plots, 2. uniform systematic plot sampling, and 3. sampling with equidistant lines. Determination of the precision of sampling with random plots was based upon equation originating in binomial distribution. In the study of uniform systematic plot sampling, use was made of an accurate stand map of a 100-hectare forest. Emphasis was laid upon sampling with equidistant lines, in which another map, covering 900 hectares, was also used. In addition, the third map was available for comparative purposes.

The dependence of the standard error of stratum-areas in systematic line and plot sampling was examined by regression analysis; the models for regression equations were derived from random sampling formulae. It appeared that the characteristics of these formulae were applicable as variables in the regression equations for systematic samples. In addition, there were found some characteristics of the distribution of the stratum which seem to influence the error in sampling with equidistant lines.

The principal results of the study are presented in the regression equations, and in some tables compiled by means of these equations. The results as regards

uniform systematic plot sampling indicate that the use of random sampling formulae, so far commonly practised, leads to considerable over-estimation of the standard error. Nonetheless, unless relatively short intervals between sample plots are used in the forest survey made on the ground, it is of advantage to study the division of the area into strata by measuring the distribution of the survey lines in these strata.

The results achieved can principally be used in two ways: for estimation of the precision in a survey already made, or to predetermine the sample size in a survey to be made. The results may be applicable to areas ranging from 100 to 1000 hectares in size, but to much larger areas as well if the pattern of configuration is of the type of stand maps in test areas. It is apparent, however, that tests in respect of larger areas are urgently needed. Moreover, in future studies attention should be paid to the correlation between survey lines.

References

BICKFORD, C. ALLEN. 1952. The sampling design used in the forest survey of the Northeast. Journal of Forestry 50: 290—293.

COCHRAN, WILLIAM G. 1963. Sampling techniques. Second edition. New York, London.

HAGBERG, ERIK. 1957. The new Swedish national forest inventory. Unasylva 11: 3—8, 28.

JEFFERS, J. N. R. 1960. Experimental design and analysis in forest research. Stockholm.

KUUSELA, KULLERVO. 1960. Variation of the site pattern and its effect on the precision of forest inventory. (In Finnish with a summary in English.) Acta Forestalia Fennica 72.

LOETSCH, F., and K. E. HALLER. 1964. Forest inventory. Volume I. München, Basel, Wien.

MATÉRN, BERTIL. 1960. Spatial variation. Meddelanden från Statens Skogsforskningsinstitut 49.

OSBORNE, JAMES G. 1942. Sampling errors of systematic and random surveys of covertype areas. Journal of American Statistical Association 27: 256—264.

SEIP, HANS KRISTIAN. 1964. Tremåling. (Forest mensuration; in Norwegian.) Skogbruksboka 3: 229—292. Oslo.

List of symbols used

A = total area

a, b, c, d = constants

AE = allowable error

K = number of samples

l = effective number of lines

$\log x$ means $\log_{10} x$

m_p = average number of lines crossing stratum p

N = size of the population

n = size of the sample

p = relative size of stratum

p_i = relative size of the stratum p in sample i

q = $1 - p$

R, r = correlation coefficient

S = standard error of estimate

s_{sa} = standard error of the stratum-area

s_{pl} = standard deviation of the length of lines in stratum p

s_{pl_1} = s_{pl} in metres

s_{pn} = standard error of the number of plots or the lengths of lines in stratum p

v = distance between plots or lines

$\bar{x}$ = arithmetic mean