Merlewood Research and Development Paper
Number 18

A METHOD OF ESTIMATING OPTIMUM SAMPLING NUMBERS WITH AN APPLICATION TO PRODUCTIVITY STUDIES ON VEGETATION

A. J. P. Gore

R. & D. 70/18

INTRODUCTION

Cochran in Snedecor (1956) cites several examples of sampling problems including a study of Vitamin A content of butter produced by creameries; a study of the protein content of wheat in the wheat fields of an area; a study of red blood cell counts in a population of men aged 20-30; a study of insect infestation of the leaves of the trees in an orchard; and a study of the number of defective teeth in third grade children in schools of a large city. He goes on to point out that in each of these examples a natural unit suggests itself, the creamery, the field of wheat, the individual man, the tree and the school. Furthermore each of these units can, and in the first three cases must, be sub-sampled instead of being measured completely.

This type of sampling is called sampling in two stages or sub-sampling. The first stage, primary sampling is the selection of the creameries, wheat fields etc. The second stage is the taking of a sub-sample of second stage units from each selected primary unit.

The natural extension of these ideas to three stage sampling and indeed to many stage sampling is obvious.

In this paper these ideas are developed in practical terms with special reference to the sampling of vegetation but the principles are common to all situations in which sampling by stages can or must be employed.

In the course of productivity studies on mixed plant communities it is frequently required to estimate the contributions by weight of the various components. The components may be individual species or recognisable parts of species. A common procedure is to clip quadrats of a size appropriate to the vegetation, Milner and Hughes (1968) refer to a number of papers which deal with this point. The investigator is then faced with the task of separating the material derived from each of the quadrats into the components he requires. He has several alternatives and it is the object of this paper to describe a method of finding the one which will provide the most information for the least effort.

METHODS

1. Field sampling

The primary units adopted in the example used here were treatment plots within blocks of a randomized two block experiment. Each block contained several such plots but only one treatment was used in this example. The reason for the choice was the purely practical one of having to sample the plots and to do this so as efficiently as possible. From each of the plots 5 quadrats were selected at random and clipped to ground level.

2. Laboratory sampling

The method suggested for taking representative samples from within each quadrat is dealt with here in detail because it is often a point of difficulty. "Quartering" is only one of several possible methods but is recorded because it is an accepted way of sub-sampling powdered or other divided materials in chemistry, particularly in the absence of a mechanical sample divider.

All the clipped vegetation from each quadrat must first be made as homogeneous as is compatible with identifying individual species or their parts in a subsequent sorting procedure. This may mean cutting the samples into lengths of about 2 inches.

The next step is to mix the material thoroughly and in the method recommended here the effectiveness of this mixing is a key factor in the reduction of variation between sub-samples.

The mixed material is quartered, one half (B) (combined opposite quarters) being set aside, the other half (A) being quartered again successively until a sample of workable size for sorting is obtained. This size is a matter for judgement but will be governed by the proportions of species occurring in small amounts. In the example 1/32nd of the quadrat was taken. The other half of the sample (B) is treated in an identical way.

The two resulting sub-samples (A) and (B) are sorted by hand into their components each of which is dried to approximately constant weight in the usual way, i.e. the loss of weight on further similar periods of drying at 105°C does not exceed 1%.

[Note: The sub-sampling procedure might preferably have consisted in taking two sub-samples randomly from a previously separated whole population of sub-samples but this is a laborious procedure and little would be gained]

3. Calculation of duplicate component weights per sample (quadrats)

This can be done manually and a worked example is provided in the Appendix. The calculations are simple but laborious. A PDP 8/I programme has been written in FORTRAN D called SBSM and is available either from the author or from the Digital Equipment Corporation User Society (DECUS) Librarian, Maynard, Massachusetts, U.S.A.

4. Analysis of variance

Data from the preceeding calculation for any one component at a time is analysed using a single classification analysis of variance procedure. Steel and Torrie (1960) p. 121-128 or Snedecor (1956) p. 266 provide examples which are relevant. A PDP 8/I programme AVSC has been written in FORTRAN D and can be used for this purpose. A copy of the programme is available from The Director, Merlewood or from the DECUS Librarian. Any similar programme or its manual equivalent will suffice of course.

The variance estimates for the between groups of samples (primary units - plots), between samples (quadrats) within groups, and between sub-samples within samples (quadrats) are illustrated in Table 1.

Table 1. Eriophorum vaginatum - Live parts only

Source	d.f.	S.S.	M.S.	MS is an estimate of
Plots (Groups of quadrats)	1	10,788.9	10,788.9	$\sigma^2 + 2\sigma_q^2 + 10\sigma_p^2$
Quadrats within plots	8	8,581.8	1,072.7	$\sigma^2 + 2\sigma_q^2$
Sub-samples within quadrats	10	2,752.8	275.3	σ^2
TOTAL	19	22,123.3		

Estimate of the variance due to quadrats =

$$S_q^2 = \frac{1,072.7 - 275.3}{2} = \frac{(S^2 + 2S_q^2) - S^2}{2}$$

Estimate of the variance due to plots =

$$S_p^2 = \frac{10,788.9 - 1,072.7}{10} = \frac{(S^2 + 2S_q^2 + 10S_p^2) - (S^2 + 2S_q^2)}{10}$$

Note: If, as occasionally happens the estimate of $\sigma^2 + 2\sigma_q^2$ is greater than $\sigma^2 + 2\sigma_q^2 + 10\sigma_p^2$ then the value of S_p^2 is taken as zero since it cannot be negative as the calculations would imply.

Having established S^2, S_q^2 and S_p^2 (estimates of the separate variances) it is possible to estimate $(S\bar{p})^2$ the variance of the plot mean, for any number of sub-samples per quadrat and number of quadrats.

$$(S\bar{p})^2 = \frac{1}{d} \left[S^2 + nS_q^2 + nmS_p^2 \right]$$

where d = k x n x m = 2 (plots) x 5 (quadrats) x 2 (sub-samples)

These can be plotted graphically as in Figure ■ to illustrate clearly the proportionate effects of different combinations of quadrat and sub-sample numbers. From this it is possible to judge the best use of the available resources. Worked examples of the individual stages are provided in the Appendix.

In the case of living Eriophorum vaginatum illustrated on the left side of Figure ■, the main source of variation is between plots and relatively little work on the component selected is necessary to approach the minimum variance. In fact three samples (quadrats) with only one, 1/32nd sub-sample would provide an estimate whose error was only 2.0% of the mean greater than a similar estimate based on 10 quadrats each sub-sampled 5 times (i.e. 50 samples or over 16 times as much work).

It must be appreciated that not all components will behave in the same way and it will depend on the requirements of the study as to the number of samples and sub-samples decided upon. For illustration, in the example cited Calluna vulgaris formed a much small proportion of the vegetation and variation about the plot mean was greater than in the Eriophorum vaginatum component. The right side of Figure ■ shows this quantitatively. For emphasis on Calluna it might be considered worth the trouble to take a single sample from each of 10 quadrats but in any case a suitable compromise between the sampling characteristics of the species present is rationally possible. Clearly, these procedures can be improved further by fully automating the calculations, but are presented at this stage as an interim report.

REFERENCES

Milner, C. and Hughes, R. Elfyn, (1968). "Methods for the Measurement of Primary Production of Grassland" IBP Handbook No. 6. Blackwell, Oxford.

Snedecor, G. W. (1956). "Statistical Methods". Iowa State University Press - Edition 5.

Steel, R. G. D. and Torrie, J. H. (1960). "Principles and Procedures of Statistics - with special reference to the Biological Sciences". McGraw-Hill Book Co. New York.

APPENDIX

1. Example* calculation of duplicate sub-sample values from primary dry weight data

Code-Plot/Quadrat/Sub-sample	Species Part	Sample wt (g)	Sub-sub-samples as % of Total (1)	Calculated wt of sub-sub-sample of Total (2) (g)	Components wt as % of Total (3)	Calculated wt of each component (g.quadrat^{-1})
I/1/(A)	E. vag. live	0.3740	21.23	2.88	15.80	100.80
	dead	1.3328	75.64	10.27	56.33	359.39
	D. flex	0.0552	3.13	0.43	2.36	15.06
	Total (1)	1.7620	100.00			
	Bulk	11.8232				
	Total (2)	13.5852		13.58		
	Moss	3.2263			17.70	112.93
	Calluna	0.2785			1.54	9.76
	Misc.	1.1410			6.27	39.94
	Total (3)	18.2310			100.00	
I/1/(B)	E. vag. live	0.3991	26.15	3.18	18.94	120.84
	dead	1.0381	68.02	8.28	49.31	314.60
	D. flex	0.0889	5.83	0.71	4.23	26.98
	Total (1)	1.5261	100.00			
	Bulk	10.6459				
	Total (2)	12.1720		12.17		
	Moss	3.1904			19.00	121.22
	Calluna	0.1400			0.84	5.31
	Misc.	1.2902			7.68	49.02
	Total (3)	16.7926			100.00	
	Bulk	602.95				
	Total (4)	638.74				

*This one of 10 similar calculations used in the examples cited on p. 2

APPENDIX (continued)

2. Example calculation of different estimates of the variance of the plot mean $(S_{\bar{p}})^2$ using the analysis of variance data given in Table 1, p. 3, together with different combinations of numbers of quadrats and of sub-samples per quadrat.

$$S^2 = 275.3 = 275.3$$

$$S_q^2 = \frac{1,072.7 - 275.3}{2} = 398.7$$

$$S_s^2 = \frac{10,788.8 - 1,072.7}{10} = 971.6$$

i. <u>1 sub-sample, 1 quadrat</u>

$$(S_{\bar{p}})^2 = \frac{1}{2} \left[275.3 + 398.7 + 971.6 \right] = 822.8$$

$$\sqrt{822.8} = 28.7$$ (or 29.6% of mean)

ii. <u>1 sub-sample, 5 quadrats</u>

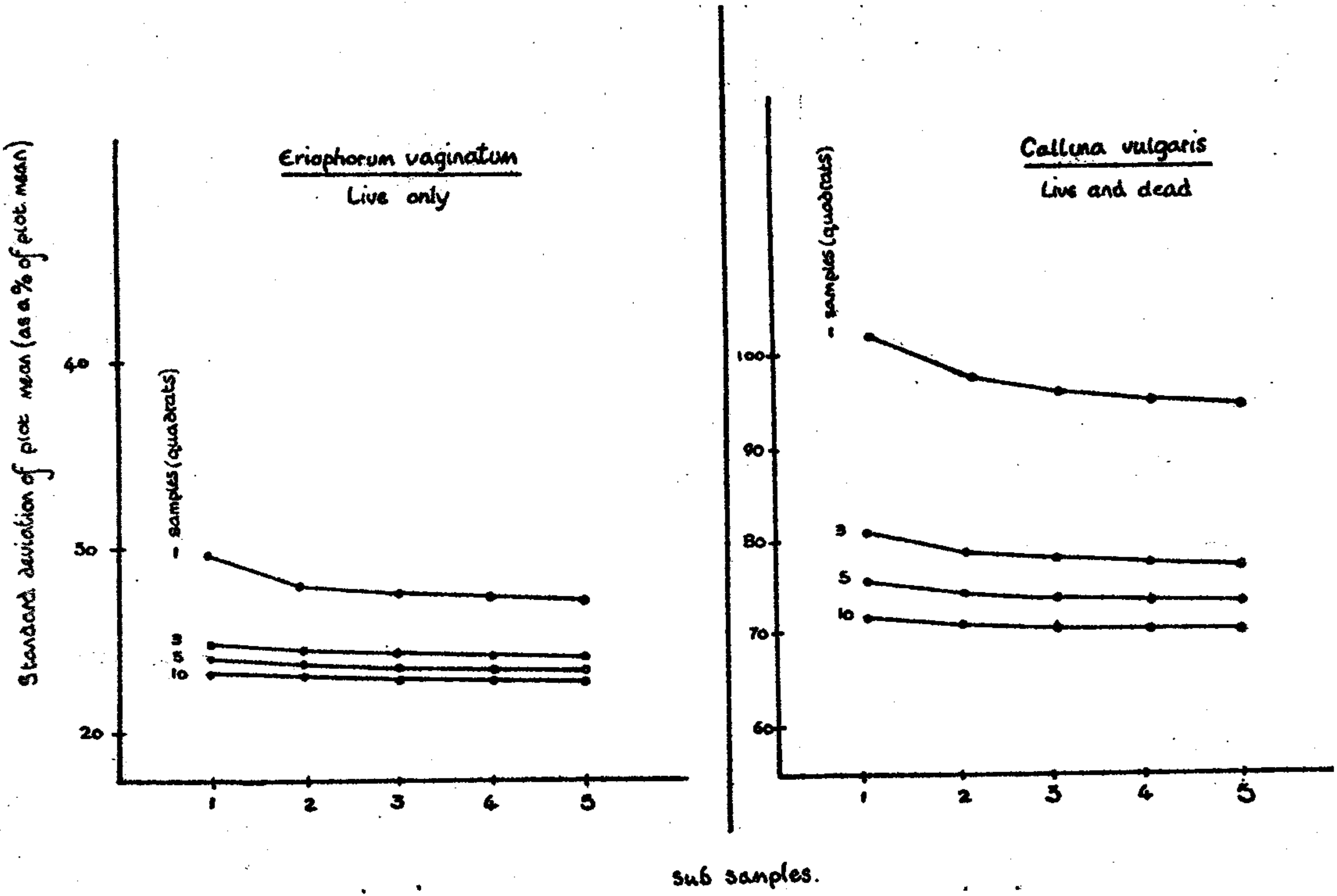
$$(S_{\bar{p}})^2 = \frac{1}{10} \left[275.3 + 398.7 + 4,858.0 \right] = 553.2$$

$$\sqrt{553.2} = 23.6$$ (or 24.3% of mean)

iii. <u>5 sub-samples, 10 quadrats</u>

$$(S_{\bar{p}})^2 = \frac{1}{100} \left[275.3 + 1,993.7 + 48,580.3 \right] = 508.5$$

$$\sqrt{508.5} = 22.6$$ (or 23.2% of mean)



Standard deviation of plot mean (as a % of plot mean)
Eriophorum vaginatum
Live only
- samples (quadrats)
1
3
5
10
20
30
40
Calluna vulgaris
Live and dead
- samples (quadrats)
1
3
5
10
60
70
80
90
100
1
2
3
4
5
sub samples.