Discussion and Development of the Point-Centred Quarter Method of Sampling Grassland Vegetation

A. HEYTING

Biometrician, Department of Research and Specialist Services, Ministry of Agriculture, Rhodesia.[1]

Highlight

The point-centered quarter method of sampling grassland vegetation is critically examined. Statistical techniques of processing data collected by the method are presented. These techniques are discussed in the light of findings from a sampling experiment at Matopos Experiment Station, Rhodesia.

Excellent methods of grassland sampling, the results of which are amenable to statistical analysis, have been developed for dense vegetation and for situations where the grass is not grazed or cut. Prominent among these are the "point methods." If the vegetation is very sparse then these methods are rather inefficient, especially if interest is centered on measurements made at ground level, as may be the case when the grassland under study is being grazed by game or cattle. In such situations, distance measurement methods, such as the point-centered quarter (P.C.Q.) method, may be useful. By these methods positive information on species composition and density is obtained at all sampling positions whereas, with the point methods, most sampling points would yield no more information than that no "hit" can be recorded.

The P.C.Q. method was developed for the determination of tree and sapling composition of woodland by Cottam and Curtis (1956) and was first adapted for application to grassland studies by Dix (1961). In this paper the method is discussed critically. Attention is confined to discovering changes in vegetation density and composition

[1] At present: Statistician, Statistics Dep., National Council for Applied Research (T.N.O.), Wageningen, Holland.

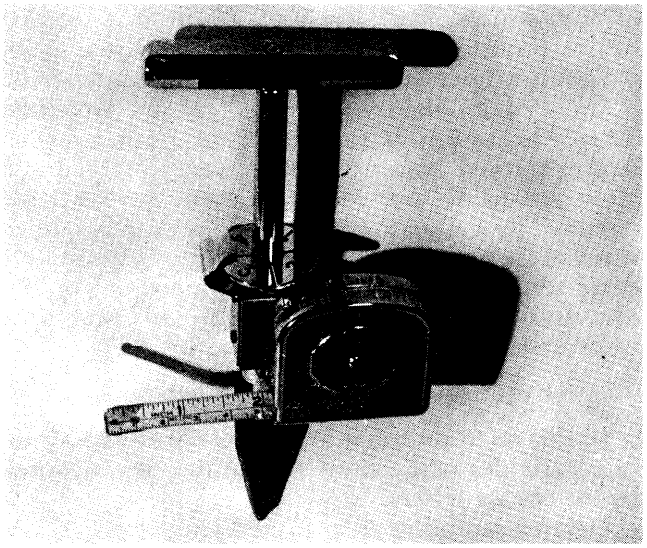

FIG. 1. The P.C.Q. calibrator.

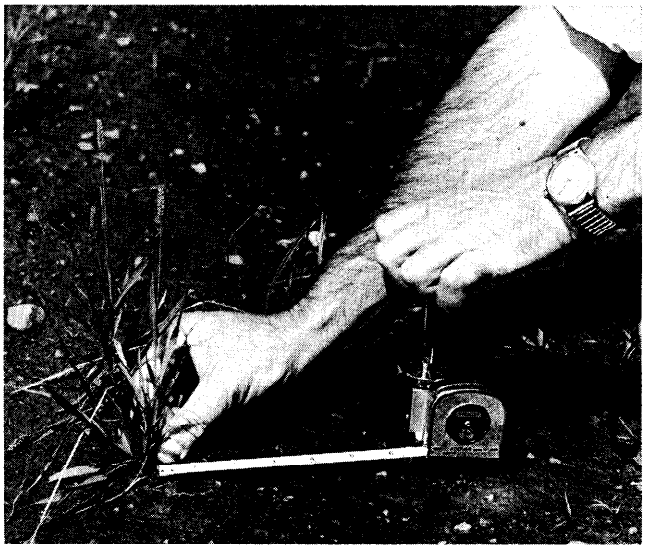

FIG. 2. The P.C.Q. calibrator in use.

while the pattern of vegetation distribution remains relatively stable. The study of pattern itself will be dealt with in a future paper. Statistical techniques for the processing of P.C.Q. data are presented and examples of their application are given.

Procedures

The below description of field procedures closely follows the directions given by Dix (1961).

Sampling points are distributed within the sampling area according to a systematic pattern. They are situated at equal intervals along parallel straight lines. Points on alternate lines are staggered. If the distance between successive points on each line is d, then the lines themselves must be a distance $\frac{1}{2}d\sqrt{3}$ apart. This produces a lattice of points such that each point is equidistant from its six nearest neighbors. In practice these instructions need not be followed rigidly, the main aim being to get a roughly uniform distribution of sampling points. The location of points is facilitated by the use of a surveyor's arrow and compass. Each sampling point is found by pacing a predetermined number of steps along a compass line and then placing the arrow vertically into the ground, guided by a notch cut in the tip of the sampler's boot.

The area around each sampling point is divided into four quarters, delineated by two lines through the sampling point, one parallel to and another perpendicular to the direction of the compass line. The demarcation of quarters is aided by filing four small marks on the surveyor's arrow in such a way that each mark is 90° from its nearest neighbor. I have chosen the name "station" to describe a sampling point together with the four quarters around it.

Within each quarter, the distance from the sampling point (base of the surveyor's arrow) to the nearest living herbaceous shoot (nearest point of root development for creeping species) is measured at ground level and recorded by species and by station.

Two minor departures from the above procedure are suggested. Firstly, the orientation of quarters need not be constant for all stations. Effort and time may be saved to a minor extent by placing the surveyor's arrow without reference to the direction of the marks which demarcate the quarters, and this is recommended. Secondly, it has been found that replacing the compass line by a steel tape or by pacing between sighting rods both work quite well in open grassland areas. Every element of choice on the part of the operator is removed if predetermined positions along the tape define sampling points. This is desirable to avoid biased findings.

Instruments.—I have developed an instrument to make the field work easier. It is called the "P.C.Q. calibrator" (Fig. 1 and 2) and is a more sophisticated version of the surveyor's arrow. The measuring device, in this instance a tape measure, is attached to the instrument in such a way that it can rotate around the central shaft. The equivalent of filing marks on the surveyor's arrow is a dial which is attached to the shaft just above the tape measure together with an indicator hand which is fastened to the tape measure itself. The tip of the shaft holds a sharpened blade which prevents the shaft from turning after it has been inserted in the ground.

Data Recording.—The data form of Fig. 3 has shown itself to be quite convenient for subsequent data processing.

SITE Paddock 8 Experiment E19 SAMPLING DATE 14/2/1966

SPECIES	STATION 1	2	3	4	5	6	7	8	9	10	Σd	n_{i1}	n_{i2}	n_{i3}	n_{i4}
H. contatus	18			2 2			13				144	4	2	1	—
		21	15	19			10		24 15 10						
D. memnalis	31 36 6					19 8		29 17 21 31			313	2	3	1	1
	44 15					16 8	19 13								
T. triandra	23				4 5						43	1	—	1	—
					11										
B. insculpta		7						25			71	2	1	—	—
		26							13						
Rl. repens			7				22				60	4	—	—	—
					6			25							
C. plurinode										20	33	2	—	—	—
				13											
TOTALS	116	78	55	31	26	51	64	92	75	76	664	15	6	3	1

FIG. 3. Data sheet for the recording of P.C.Q. observations containing fictitious records.

Except for the last five columns and the last line, the form is identical to that suggested by Dix (1961).

The stations in a single sampling area are numbered consecutively on successive data forms. The entries corresponding to stations and species are the recorded distances. Note that there are four such distances per station. The Σd column contains the total distance for each species. The n_{ij} columns contain the number of times a species is encountered j times. There are, for example, three stations at which *D. nemoralis* is encountered twice (stations 1, 8, and 9). The bottom line contains all the column totals.

Two checks have to be made. Firstly, the sum of the first 10 column totals must, of course, equal the total of the Σd column. This column is, in fact, included solely to provide this check. Secondly, if the sum of the n_{ij} column is denoted by $n_{.j}$ and the number of stations on the form is denoted by n, then the following must be true:

$$\sum_{j=1}^{4} (jn_{.j}) = 4n$$

For example, from the data sheet of Fig. 3 we have:

$$1 \times 15 + 2 \times 6 + 3 \times 3 + 4 \times 1 = 4 \times 10.$$

For the entire sampling area a summary of the last four columns is made, so that n_{ij} on the summary is the total number of stations in the sample where the ith species is recorded j times. Equating n now to the total number of stations in the sample, the second check above must, of course, apply equally to this summary.

The purpose of recording the n_{ij} and the total distances per station will become clear in subsequent sections.

Vegetation Parameters

Dix (1961) defines several parameters by which to describe the vegetation.

The relative frequency of a species is the number of stations at which the species is recorded at least once, expressed as a proportion of the total number of stations in the sample. The relative density of a species is that proportion of the total number of quarters in the sample where the species was recorded. Both these quantities may be expressed as percentages. The comparison between the relative frequency and relative density of a species gives an indication of the manner in which the species is distributed over the sampling area. A relatively high relative frequency indicates a wide distribution of small clumps while near equality of the two implies a concentration of the species in large clumps.

The mean distance per quarter is calculated and Dix claims that its square gives the mean area per shoot. This claim is based on work done by Cottam et al. (1953) who showed that if the P.C.Q. method is applied to sample a population of randomly dispersed points then the square of the mean distance per quarter provides a close estimate of the mean area per point. Unfortunately, the application of this result to the P.C.Q. sampling of grassland will only be valid in rare instances. Grassland vegetation is usually arranged in clusters of individual plants, each of which may consist of numerous shoots. In such cases the distribution of shoots cannot be assumed to be random and the mean area as obtaincd above loses its meaning. This observation is supported by statistical evidence, derived from data which were obtained at a sampling site at Matopos Experiment Station, Rhodesia. There the four distances which are measured at a single station were found to be positively correlated, thus disproving a hypothesis of random distribution of shoots. The evidence is presented in greater detail in the next section.

Some Statistical Results

In this section capital letters will be used to indicate variables while their realizations are denoted by lower case letters.

Given a sampling area, a fixed sample size, n say, and a method of selecting the n sampling points, the following notations apply:

X_{ik} the variable which records the number of quarters in which the ith species is encountered at the kth station,

p_{ijk} the probability that, for any one sample, X_{ik} has a realization equal to j,

F_i the relative density $= \left(\sum_{k=1}^{n} X_{ik}\right) \Big/ 4n$,

$E(F_i)$ the expectation of F_i, also called the "true relative density" corresponding to the sampling method,

N_{ij} the variable which counts the total number of x_{ik} in a sample which are equal to j,

D_{km} the variable which measures the distance in the mth quarter of the kth station of a sample,

$D_{k.}$ $\left(\sum_{m=1}^{4} D_{km}\right) \Big/ 4$, the mean distance per quarter at the kth station,

$D_{..}$ $\left(\sum_{k=1}^{n} D_{k.}\right) \Big/ n$, the mean distance per quarter over all stations,

$E(D_{..})$ the expectation of $D_{..}$, also called the "true mean distance per quarter" corresponding to the sampling method,

$v(F_i)$ the estimate of variance of F_i and

$v(D_{..})$ the estimate of variance of $D_{..}$.

It follows directly from the above definitions that the relative density (f_i) and the mean distance per quarter ($d_{..}$), both calculated from the sample data, are unbiased estimates of the true relative density of the ith species and the true mean distance per quarter respectively. The easiest way to calculate f_i is as

$$f_i = \left(\sum_{j=1}^{4} jn_{ij}\right) \Big/ 4n,$$

while the relation $\sum_{1} f_i = 1$ provides a check on the calculations.

Estimation of the Variance of the Relative Density

Per definition,

$$\text{variance}\,(X_{ik}) = E[\{X_{ik} - E(X_{ik})\}^2] = \sum_{j=1}^{4} (j^2 p_{ijk}) - \left(\sum_{j=1}^{4} j p_{ijk} \right)^2.$$

On the assumptions that, for fixed i, the X_{ik} are mutually uncorrelated and that, for fixed i and j, the p_{ijk} are constant and equal to p_{ij} say,

$$\text{variance}\,(F_i) = \text{variance} \left[\sum_{k=1}^{n} (X_{ik})/4n \right] = \frac{1}{16n^2} \left[\sum_{k=1}^{n} \text{variance}\,(X_{ik}) \right] = \frac{1}{16n} \left[\sum_{j=1}^{4} (j^2 p_{ij}) - \left(\sum_{j=1}^{4} j p_{ij} \right)^2 \right],$$

and p_{ij} may be estimated by $\hat{p}_{ij} = n_{ij}/n$. An estimate of variance of the relative density of the ith species may then be obtained by replacing p_{ij} in the above expression by its estimate, whence

$$v(F_i) = \frac{1}{16n^3} \left[n \sum_{j=1}^{4} (j^2 n_{ij}) - \left(\sum_{j=1}^{4} j n_{ij} \right)^2 \right],$$

which may be expanded to give

$$\begin{aligned} v(F_i) = \frac{1}{16n^3} [& n_{i1}(n - n_{i1}) + 4n_{i2}(n - n_{i2}) \\ & + 9n_{i3}(n - n_{i3}) + 16n_{i4}(n - n_{i4}) \\ & - 4n_{i1}n_{i2} - 6n_{i1}n_{i3} - 8n_{i1}n_{i4} \\ & - 12n_{i2}n_{i3} - 16n_{i2}n_{i4} - 24n_{i3}n_{i4}]. \end{aligned}$$

Variance of the Relative Density with Systematic Sampling

If the sampling points are randomly dispersed over the sampling area, the above assumptions are met and the variances of the relative densities may be estimated as shown. In practice, perfectly random distribution of sampling points cannot be achieved and even approximate randomization is very laborious. The method is to select at random intersections of the lines of a fine, imaginary grid which is superimposed on the sampling area. Sampling points can be located in the field by reference to a set of coordinate axes.

With systematic sampling there can be no guarantee that the conditions of the previous section are met. If interest is confined to the specific arrangement of plants as found in the sampling area at the time of taking the sample, this will certainly not be the case. If, on the other hand, a process is hypothesized by which the vegetation pattern is generated and if interest attaches to the vegetation considered as a random realization of this process, then there are two situations in which $v(F_i)$ computed from systematic sample data may be a valid estimate of the variance of F_i. In the first and infrequently realistic situation, the distribution of shoots is considered to be random. Under this assumption and provided that the density of sampling points is not too high, random and systematic samples are equivalent. In the second and more commonly applicable case, individual plants are arranged into distinct clumps of different species composition and density (aggregated distributions). Under the assumption that the spatial arrangement of clumps shows no systematic trends with respect to both clump size and composition, in other words that nature has provided an element of randomization, and with sampling points somewhat further apart than the diameter of the largest clumps, the results of the previous section may be applied.

There are infinitely many ways in which an aggregated distribution can fail to possess the (statistically desirable) properties which are listed in the previous paragraph and it is not possible to ascertain that a given sampling area exhibits these properties to perfection. The best that can be done is to make sure that the more likely ways in which the vegetation may differ from the statistical ideal do not occur and that sampling points are sufficiently far apart.

If the distance between adjacent sampling points in an aggregated distribution is less than the diameter of the largest clumps then this has the effect that the recording of a species at a station increases the probability of recording the same species at the neighboring stations and the failure to record the species decreases this probability. The same will be true if neighboring clumps are similar in composition, unless stations are very far apart. Occasionally it is possible to discover, for aggregated distributions, whether stations are too closely spaced. If it is hoped that individual plants may be assumed to be randomly dispersed, the same test may reveal clustering of a species.

Consider a sampling area with $n = bs$ regularly distributed sampling points, divided into b blocks, each consisting of s neighboring stations. Let k_{ih} be the observed frequency of the blocks where the ith species has been recorded h times. Clearly h may vary from zero to 4s and

$$\sum_{h=0}^{4s} k_{ih} = b, \text{ for all i.}$$

On the null hypothesis that the stations are sufficiently far apart and that the vegetation contains the required random elements, estimates $\hat{k}_{ih}$ of the expected values of the variables of which the k_{ih} are realizations may be obtained from the multinomial distribution by summation of the coefficients of those terms in the expansion of

$$b \left(\sum_{j=0}^{4} \hat{p}_{ij} t^j \right)^s,$$

for which the power of t equals h.

If stations are too closely spaced, the $\hat{k}_{ih}$ are

Table 1. χ^2 **tests to determine required spacing of sampling points for the Matopos thornveld.**

Number of shoots recorded per block	1st blocking scheme (smallest distance between points 47 inches)			2nd blocking scheme (smallest distance between points 80 inches)			3rd blocking scheme (smallest distance between points 125 inches)		
	O Observed frequency	E Expected frequency	O–E Difference	O Observed frequency	E Expected frequency	O–E Difference	O Observed frequency	E Expected frequency	O–E Difference
0	16	7.05	8.95	9	7.64	1.36	12	11.23	0.77
1									
2	9	12.44	– 3.44				7	10.50	–3.50
3	28	21.25	6.75	10	7.84	2.16	12	11.18	0.82
4	27	29.92	– 2.92	10	10.67	–0.67	11	11.34	–0.34
5	32	35.44	– 3.44	14	12.17	1.83	9	7.70	1.30
6	26	36.38	–10.38	10	12.07	–2.07			
7	32	32.66	– 0.66	5	10.43	–5.43	9	8.04	0.96
8	22	25.83	– 3.83	8	7.96	0.04			
9	21	17.97	3.03	5	5.33	–0.33			
10	9	11.01	– 2.01	9	5.87	3.13			
11–16	18	10.03	7.97						
χ^2	χ^2 (10 d.f.) = 25.80			χ^2 (4 d.f.) = 6.03			χ^2 (1 d.f.) = 1.59		
P	$P < 0.01$			$P > 0.1$			$P > 0.2$		

likely to exceed the k_{ih} for very small and for very large h and the opposite will be true for intermediate h. Disagreement with the null hypothesis may be tested by means of a χ^2 test. Before χ^2 is calculated, it must be ascertained that all expected frequencies $\hat{k}_{ih}$ achieve a minimum value of 5. If one or more of the $\hat{k}_{ih}$ are smaller than 5, successive values of h must be grouped and the corresponding k_{ih} and $\hat{k}_{ih}$ must each be pooled, until all groups comply with this requirement. A significant χ^2 value is proof that, at this spacing of points, the null hypothesis is not true. Especially if the deviations of observed from expected frequencies follow the trend described above, there is hope that increasing the distance between stations will rectify the situation.

There is a difficulty about the correct number of degrees of freedom for χ^2. If after grouping there are c frequency classes, then χ^2 has between (c – 1) and (c – 5) degrees of freedom. A significant result with (c – 1) degrees of freedom establishes significance beyond all doubt, while nonsignificance is ascertained if χ^2 with (c – 5) degrees of freedom is not significant. In other cases, there is uncertainty. A test that is both more powerful and always yields a clearcut result is being developed.

The above test was applied to data collected at Matopos Research Station, Rhodesia. The site is located in the "thornveld" which has been described by Rattray (1957). At this site, the distribution of plants is distinctly aggregated. Only perennial grasses were recorded. Nine hundred and sixty sampling points were located in 48 parallel lines, all 2 ft distant from their nearest neighbor, while the distance between points in the same line was 80 inches. Points in alternate lines were staggered.

Three blocking schemes were used and, for each of these, a different spacing of points was simulated by considering different sets of selected points only. In the first blocking scheme all points were considered, in the second all points in every third line were taken into account and in the third, every alternate point of every fourth line was used in such a way that points in alternate lines were staggered. For the first two schemes, blocks of four points were made up as compact as possible, while the blocks of the third scheme contained two points, one from each of two adjoining lines.

The results for *Themeda triandra*, the species with the highest relative density, are given in Table 1. The other major species gave similar results. The grouping of neighboring frequencies in the table serves to achieve the required minimum expected frequency of 5.

It appears that, for this type of vegetation, stations should not be as closely spaced as in the first blocking scheme.

A second common deviation from the statistical ideal is that a given species occurs more commonly in one part of the sampling area than in another. The χ^2 test is again suitable to detect this. The sampling area is divided into a few large patches of equal size, u in number say. Let r_{is} be the total number of times that the ith species has been recorded in the sth patch. On the null hypothesis that the distribution of the species is uniform, the expected frequencies are all equal and may be estimated as $\hat{r}_i$, where

$$\hat{r}_i = \left(\sum_{s=1}^{u} r_{is} \right) \Big/ u.$$

Table 2. Analysis of variance of distance measurements (the unit in the analysis is d_{km} in cm).

Source of variation	Selection (1) (smallest distance between points 7 ft) D.F.	Mean Squares	F and Signif.[1]	Selection (2) (smallest distance between points 13 ft) D.F.	Mean Squares	F and Signif.	Selection (3) (smallest distance between points 20 ft) D.F.	Mean Squares	F and Signif.
Sets	—	—	—	—	—	—	2	104.26	< 1
Blocks (within sets)	79	268.38	1.77**	19	294.58	1.96*	6	106.72	0.71 (NS)
Stations within blocks	240	152.03		60	150.30		27	150.58	2.26**
Quarters within stations	—	—	—	—	—	—	108	66.52	

[1] * and ** signify significant at $P \leqslant 0.05$ and $P \leqslant 0.01$ respectively. (NS) means nonsignificant.

Chi-squared tests the disagreement between the r_{is} and $\hat{r}_i$. The number of degrees of freedom for χ^2 is $(u-1)$. If the probability level corresponding to χ^2 is small then there is strong evidence that the distribution of the species is heterogeneous. The cure is then to divide the sampling area into strata which, regards the distribution of species, are uniform within themselves, to calculate f_i and $v(F_i)$ for each stratum and, possibly, to pool these quantities over all strata.

This test was applied to the Matopos data. The sampling area was divided into four patches, each 48 ft × 66 ft 8 inches in size. For *Themeda triandra* and using the points of blocking scheme 3, the test yielded χ^2 (3 degrees of freedom) = 3.33; $0.3 < P < 0.5$. Similar results were obtained for the other major species. There is thus no evidence that the species distribution on this site is not homogeneous.

While these two tests cope with what are thought to be the more common situations in which care must be exercised with the statistical evaluation of P.C.Q. data from systematic samples, many more may be listed. Examples are systematic variation in clump size together with a tendency for some species to occupy specific positions in the clumps and periodicity in the distribution of species. Operators should always be on the lookout for like trends.

It is not recommended that both χ^2 tests are always applied for all species. Once an operator is familiar with a certain vegetation type, he will be able to judge adequately what distances to use between points and close scrutiny of the data will reveal marked heterogeneity of distributions of species without recourse to formal tests of significance. It is perhaps reassuring that, by arguments similar to those of Yates (1949), $v(F_i)$ computed from systematic sample data, if at all biased, is likely to be an over-estimate of the variance of F_i.

Distribution of the Mean Distance with Systematic Sampling

By arguments similar to those of the previous section, the $D_{k.}$ corresponding to a systematic sampling method may be uncorrelated, in which case the estimate of variance of $D_{..}$ is the same as for a random sample, nl.

$$v(D_{..}) = \left[\sum_{k=1}^{n} (d_{k.} - d_{..})^2\right] \Big/ (n-1)n.$$

In accordance with the discussion of intraclass correlation by Scheffé (1959), correlation between the $D_{k.}$ may be discovered by means of an analysis of variance. The procedure is illustrated for the Matopos data. Different distances between sampling points were simulated by using only selected points. The selections used were, (1) all points of every third line, (2) alternate points of every sixth line, and (3) every third point of every ninth line. The corresponding minimum distances between sampling points are 7, 13, and 20 ft respectively. For each selection, square blocks of four adjacent points were made up and the analyses of variance of Table 2 were obtained.

The F value in the "blocks" line of the analysis is the ratio of the mean squares for blocks and stations within blocks and tests whether the mean distances at stations in the same blocks are correlated. If the blocks mean square is the larger of the two, this indicates positive correlation.

For selection 3, the analysis was extended somewhat. There was a suggestion that the distribution of $D_{k.}$ was not homogeneous over the sampling area. The blocks were therefore grouped into sets of three blocks each, roughly corresponding to areas that were suspected to have different densities. The sum of squares for blocks was partitioned into the sum of squares for sets and the sum of squares for blocks within sets. If the mean square for sets exceeds the mean square for blocks within sets, then the F ratio of these two mean squares provides a test of disagreement with the hypothesis that $E(D_{k.})$ is constant over the entire sampling area. As it happens, no evidence was found that the spatial distribution of mean distances at the Matopos site is heterogeneous. If heterogeneity had been found, the position could have been saved by stratification in the way described for relative densities.

Also for selection 3, the four distance measurements at each station were included in the analysis. The F value in the "stations within blocks" line is the ratio of the last two mean squares. If the mean square for stations within blocks is the larger, as in this case, a positive correlation between the distances measured at the same station is indicated.

The results for the Matopos site are much as would generally be expected. Close spacing of points leads to a positive correlation between the mean distances at neighboring stations, which disappears as stations are further apart. At Matopos, all evidence of correlation between the $D_{k.}$ disappears at a spacing of 20 ft (selection 3). The positive correlation between the four distances measured at the same station, which was found at Matopos, may be expected to crop up in most types of grassland. This is the reason for estimating the variance of $D_{..}$ from the mean distances per station instead of the distances per quarter. Computationally, it is easiest to replace the $d_{k.}$ in the formula for $v(D_{..})$ by the total distances per station and $d_{..}$ by their mean and to divide the answer by 16.

The analysis of variance procedure is strictly valid only if the $D_{k.}$ are normally distributed, in which case absence of correlation may be taken to indicate independence of the $D_{k.}$. In very dense cover the distribution of the mean distance per station is likely to be positively skew, tending to symmetry as the vegetation gets less dense. Fortunately, the P.C.Q. method is most useful in places with sparse vegetation. Also, the analysis of variance is reasonably "robust" as far as mild deviations from normality are concerned. In cases of doubt, a cumulative frequency diagram should be plotted on normal probability paper. The Matopos data agreed reasonably well with the assumption of normality.

Considering all the evidence together, it appears that, for the Matopos thornveld, valid estimates of the variances of the F_i and $D_{..}$ may be obtained from a systematic sample, provided that sampling points are about 20 ft or more apart.

Statistical Analysis

Because of the many designs of experiments and surveys which may yield P.C.Q. data, it is not possible to present a general method of statistical analysis. The best that can be done is to mention a few of the considerations and techniques which can play a part in the construction of a suitable method for a given set of data. Attention has been confined to the univariate case, i.e. the analysis of the relative density of one species at a time.

With standard designs, involving adequate replication, it is usually possible to employ analysis of variance techniques with the relative density of a species as variable. Similarly, analysis of covariance may be done, with as covariable for example the relative density of the species at the initiation of the experimental treatments. Two necessary conditions for a valid analysis of variance are that the variable is normally distributed and that its variance is constant. As will be seen later, the normality assumption will often be met reasonably well. Difficulties are likely to arise with the second condition.

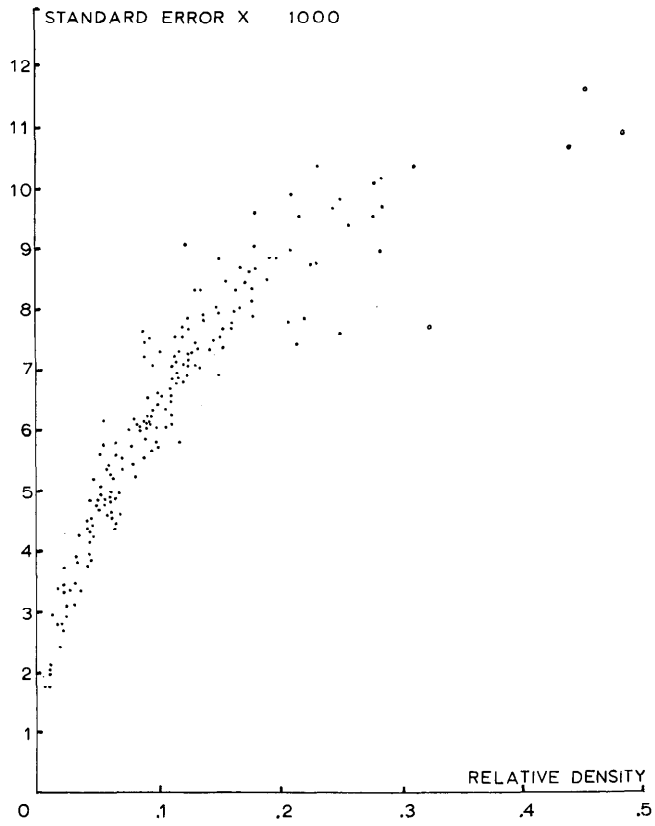

FIG. 4. Relationship between relative density and its standard error. Data from four Matopos grazing experiments.

The analysis of four grazing experiments at Matopos necessitated the computation of the $v(F_i)$ corresponding to 192 relative densities, each based on observations at 1,000 stations. Fig. 4 shows the observed relationship between these f_i and $v(F_i)$ (actually, $1000[v(F_i)]^{\frac{1}{2}}$ was plotted). No data were available for f_i greater than 0.5. The obvious similarity with the behavior of binomial data is not surprising and may be expected to occur in general.

The residual variance of the relative density per plot in an analysis of variance consists of a component due to within plot variation, estimated by $v(F_i)$, and a component due to variation between plots. If the true relative densities of a species, corresponding to different treatments, vary considerably, the first component will not be constant and this is likely to cause heterogeneity of variance.

This is not likely to be very important in practice, unless treatment differences approach the order of obviousness. Although there may then be little interest in tests of significance, confidence intervals remain important and a suitable variance-stabilizing transformation should be applied. Care must be taken that the normality assumption is not violated over-much after such a transformation. It is unfortunate that, in contrast to binomial variables, there is no known, theoretically justified transformation. Furthermore, the relationship of Fig. 4 may well vary with the type of vegetation. In some cases, distribution-free methods may be useful.

In practice, considerations of space and/or time will often severely limit the level of replication. By making use of the $v(F_i)$, it is possible to devise methods for the construction of confidence intervals for the relative densities corresponding to the different treatments in an unreplicated experiment and to test for treatment differences. The interpretation of the results of such an analysis is complicated by the fact that treatment effects and "plot" effects are completely confounded. While every care can be exercised to minimize plot effects (choice of a uniform experimental area; careful management; selection of a suitable variable for analysis, such as the difference between the relative density of a species at the beginning and at the end of the experiment), there always remains a component of the variability between plots which is beyond the experimenter's control. Before the methods which are described below can be employed, there must be some reassurance that this component is relatively small.

One commonly encountered unreplicated situation arises when it is desired to investigate a time trend at a single site under constant management. Here the difficulty is not one of plot effects but of confounding between time and seasonal effects. Replication can never help to overcome this hurdle and neither does staggering of the starting times eliminate all difficulties. All the same, it may be interesting to study the combined effect of time and season, in which case the following methods can be applied.

From the definition of F_i,

$$F_i = \left(\sum_{k=1}^{n} X_{ik}\right) \Big/ 4n,$$

and the Central Limit Theorem of probability theory, it follows that F_i is asymptotically normally distributed with increasing n. Provided that n is large and that the standard error of f_i is small compared with both f_i itself and the quantity $(1 - f_i)$, statistical treatment of F_i as a normally distributed variable with known variance equal to $v(F_i)$ should give a good approximation. As will be seen later, n will often need to be very large, in order to achieve the desired degree of precision.

There is an obvious analogy between the statistical methods which are applicable in an experimental situation where the members of a (sub)set of q treatment means are based on different numbers of replications, $r_1, r_2, \ldots, r_q$, say and those that can be used when dealing with a set of normally and independently distributed variables, $Y_1, Y_2, \ldots, Y_q$, with unknown means, $\mu_1, \mu_2, \ldots, \mu_q$, and known variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_q^2$. Let the error variance for the hypothetical experiment be σ^2, estimated with f degrees of freedom, and assume that the residuals in the underlying mathematical model are normally and independently distributed. If now the r_i are such that $r_1\sigma_1^2 = r_2\sigma_2^2 = \ldots = r_q\sigma_q^2 = \sigma^2$, then, as f tends to inifinity, the distributions of the means in the experiment tend to the distributions of the corresponding Y_i. Thus confidence regions for functions of the μ_i can be constructed and tests of significance of such functions can be performed as for the experiment, merely replacing the r_i by σ^2/σ_i^2 and taking the degrees of freedom for error to be infinitely many. The unknown constant σ^2 plays no important role, since it is usually equally represented in the test statistic and in the variable of which the test statistic is hypothesized to be a realization. In practice therefore, the r_i can be replaced by $w_i = 1/\sigma_i^2$, the "weights" of the Y_i.

The above trick is never essential for the derivation of a result. It may, however, provide a useful shortcut. Identifying the Y_i with relative densities and the σ_i^2 with the corresponding $v(F_i)$, the following are some results with obvious applications to P.C.Q. problems.

(I) On the assumption that $\mu_1 = \mu_2 = \ldots = \mu_q = \mu$, the weighted mean,

$$u = \left[\sum_{i=1}^{q} (w_i y_i)\right] \Big/ \left(\sum_{i=1}^{q} w_i\right),$$

is the best (minimum variance) linear unbiased estimate of μ and u may be seen as a realization of a normally distributed variable with variance equal to

$$1 \Big/ \left(\sum_{i=1}^{q} w_i\right).$$

(II) The weighted sum of squared deviations from u,

$$\sum_{i=1}^{q} [w_i(y_i - u)^2] = \sum_{i=1}^{q} (w_i y_i^2) - u^2\left(\sum_{i=1}^{q} w_i\right),$$

is a realization of a noncentral χ^2 variable with $(q - 1)$ degrees of freedom and noncentrality parameter equal to

$$\left[\sum_{i=1}^{q} \sum_{j<i} w_i w_j (\mu_i - \mu_j)^2\right] \Big/ \left(\sum_{i=1}^{q} w_i\right),$$

so that the null hypothesis $\mu_1 = \mu_2 = \ldots = \mu_q$ may be tested against the class of all alternative hypotheses by means of a χ^2 test.

(III) The above sum of squares may be partitioned orthogonally into several (up to $(q-1)$) components, to test selected hypotheses, analogous to the partitioning of the treatment sum of squares of an experiment with unequal replication (e.g. fitting polynomials of increasing degree).

(IV) A contrast γ among the means is defined as a linear function of the μ_i,

$$\gamma = \sum_{i=1}^{q} (a_i \mu_i) ,$$

with known constant coefficients subject to the condition

$$\sum_{i=1}^{q} a_i = 0 .$$

An unbiased estimate of γ is provided by

$$c = \sum_{i=1}^{q} (a_i y_i) ,$$

and c is a realization of a normally distributed variable C with variance equal to

$$\sigma^2(c) = \sum_{i=1}^{q} (a_i^2 \sigma_i^2) .$$

The quantity $C^2/\sigma^2(c)$ has non central χ^2 distribution with one degree of freedom and non centrality parameter equal to $\gamma^2/\sigma^2(c)$, so that the null hypothesis $\gamma = 0$ may be tested against the alternative $\gamma \neq 0$ by means of a χ^2 test.

(V) If the contrasts of interest are such that the corresponding tests of (IV) don't fit into the scheme described in (III), some workers may prefer to use the S method of Scheffé (1959). Adapted to the P.C.Q. problem, the procedure is to declare a contrast significantly different from zero or not at the level α, according as the interval $(c - s\sigma(c), c + s\sigma(c))$ excludes, respectively includes the value zero. The constant s is calculated from $s^2 = \chi^2\alpha;(q-1)$, where $\chi^2\alpha;(q-1)$ is the upper α point of the central χ^2 distribution with $(q-1)$ degrees of freedom. If the only contrasts of interest are pairwise differences between the expectations, a more powerful testing procedure may be obtainable by adapting one of the existing procedures for testing the differences between the means in an experiment with unequal replication. An example of such a method is the multiple range test of Duncan (1957).

Mutatis mutandis, the analysis of distance measurements is similar to that of relative densities. Provided that n is large, equating $v(D_{..})$ to the true population variance should give good approximations.

If the variables in an analysis correspond to successive samples, collected in the same sampling area, care must be exercised that these samples are independent. If random sampling is used, a fresh randomization is required for each sample. With systematic placement of points, two successive sets of observations at the same site will not be strictly independent, unless the distance between points is increased a lot. The points at the second sampling can then be fitted between those of the first, at a sufficient distance for the two samples to be independent.

Random Line Sampling

If the sampling area is too small to accommodate the required number of points at a sufficiently large distance apart and if the randomization of points over the sampling area is too laborious, then random line sampling can be used. Sampling positions are chosen at random along a number of parallel lines, like the compass lines of the systematic sample, and a new randomization is used at each sampling time. Stratification by lines or sections of lines is easily achieved. The selected sampling positions are located on the ground by means of a tape-measure.

Interpretation of Results

Grassland vegetation may change in any combination of the following two ways:

(I) the relative preponderance of all species, in terms of numbers of shoots and plants per unit area, remains constant but the overall cover alters and

(II) the overall cover, in terms of total number of shoots and plants per unit area over all species, remains constant but the relative preponderance of one or more species changes.

A significant increase or decrease in $d_{..}$ with time may reasonably be taken to mean that (I) has taken place, while a significant change in one or more of the f_i would indicate the occurrence of (II). On this basis, four grazing experiments at Matopos were analyzed, using the statistical techniques described above. The results will, in due course, be published elsewhere but it is perhaps relevant to quote the officer in charge of these experiments who wrote:

> "I am very glad to see that the P.C.Q. system seems to be giving results—at any rate the calculated results confirm trends that we have observed visually. This is something we have never been able to do with the point system."

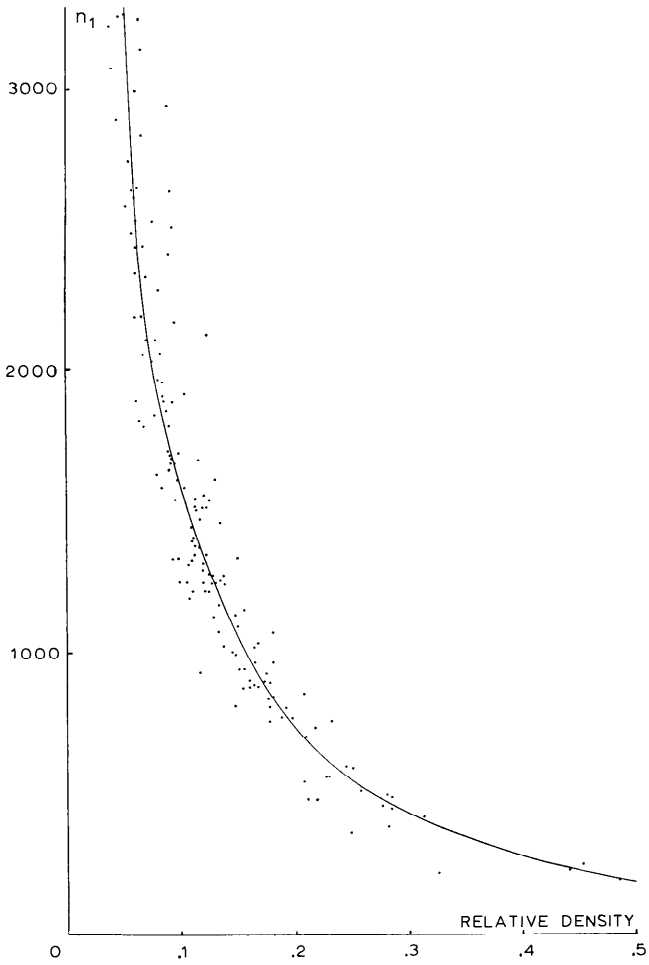

FIG. 5. Relationship between relative density and n_1, the number of points for 10% precision.

Number of Stations Required

From each pair of values $(f_i, v(F_i))$, obtained from a sample size n, it is an easy matter to obtain an estimate of n_p, the average number of sampling positions required to estimate $E(F_i)$ with such a degree of precision that the 95% fiducial limits are given by $f_i \pm pf_i/10$. It may be readily verified that this estimate is provided by

$$n_p = nv(F_i)\left(\frac{19.6}{pf_i}\right)^2.$$

The n_1 have been computed for each of the 192 sets of values of n, f_i, and $v(F_i)$ which were obtained from the Matopos grazing experiments. Fig. 5 gives the relationship between n_1 and relative density.

Insufficient is known about possible relationships between relative density and its standard error to be able to postulate a law defining the corresponding relationship between n_1 and relative density. It is quite possible that this will vary a little with locality. Despite these objections, an attempt was made to fit a curve to the points of Fig. 5. The only restrictions imposed are that it must be asymptotic to the n_1 axis, it must pass through the point $n_1 = 0$, relative density = 1, it must be continuous and without points of inflection. Of all the many types of curves tried, the best fit was achieved by

$$n_1 = 188.9\left(\frac{1-f_i}{f_i}\right)^{0.9631}.$$

For different relative densities, the n_p values may be estimated from this equation by dividing the corresponding n_1 by p^2. An estimate of p with n′ stations is provided by $(n_1/n')^{\frac{1}{2}}$.

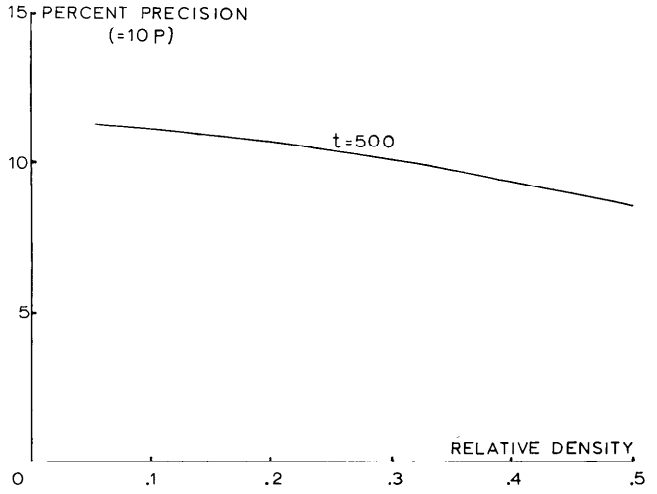

FIG. 6. Relationship between relative density and percentage precision for a fixed number of strikes (t).

Another way to present these results is in terms of the precision achieved if sampling is continued until a species of interest has been recorded a given number of times. If this total number of "strikes" is denoted by t and provided t is not too small, then for any corresponding f_i, p may be estimated from

$$p = 2(n_1 f_i/t)^{\frac{1}{2}},$$

where n_1 is obtained from Fig. 5.

Dix (1961) claims that quite good precision is achieved for t = 30. For the Matopos data, Fig. 6 gives the relationship between f_i and p for t = 500. For other values of t the corresponding p values may be found by multiplying the values of Fig. 6 by $(500/t)^{\frac{1}{2}}$. Thus the precision for t = 30 is obtained by multiplication by 4.08.

The above relations are all useful if only a rough estimate or guess of the true relative density is available and an estimate is sought of n_p for given p or of p for given n. The procedure is to replace f_i in the expressions and graphs by the rough estimate. If much sampling is done in a particular vegetation type, similar relations and graphs are

easily constructed for that specific type and may be readily updated from time to time.

In order to achieve the desired degree of precision if $E(F_i)$ is small, many points are needed. If much interest exists in rare species, then multiphase sampling, as described by Yates (1949), is recommended. In each phase, all but a number of selected grass species are ignored. By this strategem, the relative density of a rare species is artificially increased within its sampling phase.

The required number of sampling positions to estimate mean distances with sufficient precision will usually be smaller than the number needed for precise estimates of relative densities.

Acknowledgments

Thanks are due to the Director, Dep. of Research and Specialist Services, Ministry of Agriculture, Rhodesia, for permission to publish this article. The assistance of Mrs. G. Gough, Mrs. L. Murray, Mr. N. H. Judge, and Mr. D. White with the data processing and preparation of the manuscript and of Mr. W. B. Cleghorn, Mr. T. C. D. Kennan, and Mr. R. P. Denny with the collection of data in the field, is gratefully acknowledged.

LITERATURE CITED

COTTAM, G., AND J. T. CURTIS. 1956. The use of distance measures in phytosociological sampling. Ecology 37:451–460.

COTTAM, G., J. T. CURTIS, AND B. W. HALE. 1953. Some sampling characteristics of a population of randomly dispersed individuals. Ecology 34:741–757.

DIX, R. L. 1961. An application of the point-centred quarter method to the sampling of grassland vegetation. J. Range Manage. 14:63–69.

DUNCAN, D. B. 1957. Multiple range tests for correlated and heteroscedastic means. Biometrics 13:164–176.

RATTRAY, J. M. 1957. The grasses and grass associations of Southern Rhodesia. Rhodesia Agric. J. 54:197–233.

SCHEFFÉ, H. 1959. The analysis of variance. Ch. 7. Wiley & Sons, London.

YATES, F. 1949. Sampling methods for censuses and surveys. Ch. 3, 7. Charles Griffin & Co., London.