A STATISTICAL ANALYSIS OF DESERT ROCK DISTRIBUTIONS

by

Eugene Richard Heyman

A Thesis Submitted to the Faculty of the

COMMITTEE ON STATISTICS

In Partial Fulfillment of the Requirements
For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 7 2

STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Eugene R. Heyman

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

R. O. Kuehl
R. O. KUEHL
Professor of Statistics

May 1 1972
Date

ACKNOWLEDGMENTS

I wish to express my most sincere gratitude to Dr. Robert O. Kuehl for his kind help and guidance in preparing this paper. I also wish to offer my appreciation to Drs. James R. Gebert and John L. Denny for serving on my committee, to Mrs. Janet L. Beauchamp for her programming and to Dr. John L. Thames whose idea it was to study desert pavement coverage.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES--Continued

LIST OF ILLUSTRATIONS

ABSTRACT

A uniformity trial study, performed on two different six meter square sites, showed a half meter square quadrat to be the optimum size sampling unit in which desert rocks may be sampled. A nested sample was then taken from a typical Southern Arizona desert region. The data were taken as the observed frequencies of rocks in various size classes contained in each quadrat. The sample data and the uniformity trial data were fit to known theoretical distributions. Rocks 45 mm and larger were found to fit contagious distributions which showed that clumping existed among these rocks. The proportion of ground covered by rocks larger than 11.2 mm was estimated as 15.1%.

The optimum sampling scheme appeared to be a stratified sample, since the ground coverage was divided into definite strata. Exact quantification of the strata was unable to be done, however. Therefore a simple random sample seemed to be the best approach as its variance was considerably lower than that of the nested sample.

CHAPTER I

INTRODUCTION

Desert pavement coverage plays an important role in the hydrologic behavior in a desert watershed. Coverage influences surface roughness, thereby affecting the path length of water flow. It can also intercept rainfull and influence infiltration in some undetermined manner, such as channelizing water which may tend to increase erosion. The coverage could also protect the surface from rainfall, thus reducing erosion. Hydrologic models developed for desert regions should, therefore, take into account the influence of desert pavement.

Pavement coverage also influences biological factors such as vapor movement in desert regions. Vapor movement is influenced by the surface temperatures of the pavement which in turn is affected by the rocks covering it. Vapor barriers are often formed beneath rocks due to condensation and offer protection for lower forms of desert life; so deserve consideration.

For these reasons it was desirable to study the behavior of the rocks covering desert pavements. The few ground coverage studies previously done have dealt with non-desert regions, taking a descriptive, rather than a quantitative approach. Thus, it is the purpose of this study to design a procedure in which rocks may be sampled, and then to quantify some of their distribution properties.

CHAPTER II

LITERATURE REVIEW

The initial stage in sampling studies is the determination of the sampling unit size. To determine the optimum size of a sampling unit one could sample an area many times, each time using a different size sampling unit or quadrat. Greig-Smith (1964) suggested the use of a grid of small unit cells. By combining adjacent cells, quadrats of varying sizes may be formed. His method, though, entails complete enumeration of the sampling area, which becomes prohibitive in sampling a large desert region.

The most common approach to optimum sampling unit size in agricultural crop yields is a uniformity trial study, proposed by Smith (1938). He suggested the relationship,

$$V_x = V_1/x^b \qquad (1)$$

where V_x is the variance per unit area among quadrats x units in size and b is referred to as an index of soil heterogeneity. This heterogeneity index was assumed to take on values from zero to one by Smith (1938), with no further commentary. Brim and Mason (1959) gave a brief explanation, saying that a value of zero represents perfect correlation between samples or complete uniformity, and unity no correlation. They further claimed that the extreme value, $b = 1$, could develop from sites which range from complete uniformity to extreme heterogeneity. No reference has been found in the literature for values beyond one.

The logarithmic expression of the variance relationship in equation (1) is

$$\log V_x = \log V_1 - b \log x \tag{2}$$

which implies a linear regression of the log of the variance, V_x, on the log of the quadrat size, x, with the heterogeneity index as the regression coefficient.

Smith (1938) further suggested the cost function

$$\text{Cost} = K_1 + K_2 x \tag{3}$$

where K_1 is proportional to the time required to locate a sample and K_2 is proportional to the time necessary to count the rocks in one basic unit. He also showed how the cost per unit of information is a minimum when

$$x = \frac{bK_1}{(1 - b)K_2} \ . \tag{4}$$

When sampling over a large area, Cochran (1963) suggested a stratified sample if the target population is divided into distinct natural strata. He also discussed cluster sampling, which can be employed by artificially subdividing an area into equal sections. Field cluster sampling is most often referred to as nested sampling. Marcuse (1949) discussed optimum allocation for a nested sample using variance component estimates. A nested sample, according to Marcuse (1949), may be represented as

$$x_{ij} = \mu + \lambda_i + \delta_{ij} \tag{5}$$

which is the j'th subsample from the i'th sample. The value μ is the general population mean and is a fixed constant. The components λ_i and δ_{ij} are random variables, each having a zero mean and zero covariance and variances σ_1^2 and σ_2^2 respectively, called variance components. If a nested sample is used the variance of its mean is then a function of these variance components, and is

$$V = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_1 n_2} \tag{6}$$

where n_1 and n_2 are the number of samples and subsamples per sample respectively. Marcuse (1949) then gave the optimum allocations for n_2 and n_1 as

$$n_2 = \sigma_2/\sigma_1\sqrt{c_1/c_2} \tag{7}$$

$$n_1 = \frac{\sigma_1 C}{\sigma_1 c_1 + \sigma_2\sqrt{c_1 c_2}} \tag{8}$$

where C, c_1 and c_2 are from the cost function

$$C = c_1 n_1 + c_2 n_1 n_2 \ . \tag{9}$$

Equation (8) gives the optimum number of samples using a fixed cost. Bancroft and Brindley (1956) offered an optimum value for n_1 using a preassigned variance rather than a fixed cost as

$$n_1 = \frac{\sigma_1^2 + \sigma_2^2/n_2}{V} \tag{10}$$

with V being the desired variance and n_2 the optimum value from equation (7).

To statistically quantify sample data, the mean and variance of rocks per unit area are of interest. Distribution properties and the possible fitting of theoretical distributions would further describe the data. Greig-Smith (1964) discussed density and cover as two means of quantifying plants. Density is a mean number per unit area and cover is the proportion of ground covered by the plants.

Discussions concerning spatial patterns by ecologists most often begin by seeking randomness. Pielou (1969) showed that if individuals are arranged independently and randomly, then the probability that one particular unit contains x individuals is given by

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots \qquad (11)$$

which is the Poisson density with λ as the mean number of individuals per unit. She explained this by assuming every unit to be subdivided into a large number, n, of locations, none of which may contain more than one individual. Each of these locations would then have the same small probability, p, of containing an individual. The probability of a unit having exactly x locations containing an individual would then be given by the binomial probability function as

$$f(x) = \binom{n}{x} p^x (1 - p)^{n-x}, \; x = 0, 1, 2, \ldots . \qquad (12)$$

Assuming n to be large, p very small, the mean, np, of moderate size and x negligible relative to n, she demonstrated that setting $np = \lambda$, equation (12) has the Poisson as its limiting distribution for sufficiently large n.

As is often the case, actual data may show an excessive number of samples containing zero individuals and too few containing only one, causing the data to fail to fit the Poisson. Neyman (1939) called this a contagious distribution because, as he explained, one individual in a sampling unit increases the probability of more also being observed. The most common of these is the negative binomial, which Pielou (1969) showed can be generalized from a compound Poisson distribution. She gave this probability function as

$$f(x) = \frac{\Gamma(k + x)}{\Gamma(k)k^x} \cdot \frac{m^x}{x!} \left(1 + \frac{m}{k}\right)^{-k+x}, \quad x = 0, 1, 2, \ldots . \tag{13}$$

The negative bionomial has been offered as a more simplified function by Bliss and Fisher (1953) as

$$f(x) = \frac{(k + x - 1)!\ R^x}{x!\ (k - 1)!\ q^k}, \quad x = 0, 1, 2, \ldots \tag{14}$$

where R and q are functions of its two parameters, m and k, with $p = m/k$, $q = 1 + p$ and $R = p/q$. Bliss and Fisher (1953) presented three methods of estimating the parameter k, moments, number of zeroes and maximum likelihood. At least 90% efficiency may be achieved using the method of moments if $k/\bar{x} > 6$ for small values of $\bar{x}$, k exceeds 13 for large $\bar{x}$ or $(k + \bar{x})(k + 2)/\bar{x} \geq 15$ for intermediate values of $\bar{x}$. If the conditions for the method of moments are not met and at least 1/3 of the units are in the zero class, they suggested the number of zeroes method. This is an iterative solution for the estimate k from

$$k \log (1 + \bar{x}/k) = \log (N/f_0) \qquad (15)$$

where N is the sample size and f_0 is the frequency observed in the zero class. The maximum likelihood method, being reserved for cases not meeting the restrictions of the previous methods, results in an iterative solution for k from

$$\sum \left(\frac{A_x}{k + x}\right) = N \ \ln \left(1 + \frac{\bar{x}}{k}\right) \qquad (16)$$

where A_x is the accumulated frequency in all units containing more than x individuals.

Pielou (1969) stated the defining property of a contagious discrete distribution to be the variance being greater than the mean, in contrast to their being equal, as in the Poisson. She therefore suggested the variance to mean ratio as a measure of aggregation. David and Moore (1954) proposed what they called an "index of clumping" as

$$I = V/m - 1 \qquad (17)$$

where V and m are the sample variance and mean respectively. As a means of comparing two values of I they suggested evaluating

$$w = -\tfrac{1}{2} \ln \left(\frac{V_1/m_1}{V_2/m_2}\right) \qquad (18)$$

for which they derived its limiting distribution. For a .05 level of significance two values of I differ if w lies outside the range of $-2.5/\sqrt{N - 1}$ and $2.5/\sqrt{N - 1}$.

Data in this paper will most often be tested to fit either the Poisson or negative binomial distributions. If data tends toward a unimodal symmetric function, the normal approximation may well explain the data. Otherwise, the discrete distributions, Poisson and negative binomial, should adequately explain the sample data.

CHAPTER III

MATERIALS AND METHODS

Uniformity Trial

The study area was a 300m x 300m section of desert at the Silverbell site near Marana, Arizona, operated by the International Biological Program at The University of Arizona under special use permit from the Bureau of Land Management.

For the uniformity trial study, two sites six meters square were selected. It was necessary to find areas of this size without vegetation, so they were not randomly selected. When two such sites were located, each was partitioned into 144 half meter square basic units. Rocks were then enumerated in each basic unit and tabulated as frequencies in each size class and were then classified by the Udden-Wentworth scale of particle size classes from Lane (1957). As shown in Table 1, this is a logarithmic scale in which the minimum dimension of each class is twice that of the previous class.

To estimate the heterogeneity index, b, from equation (2), Smith (1938) suggested a weighted regression, whereby the weights were the respective degrees of freedom associated with each variance. This weighted regression

Table 1. Udden-Wentworth Scale of Particle Size Classes

Class Number	Dimensions
1	11.2 mm - 22.4 mm
2	22.4 mm - 45 mm
3	45 mm - 90 mm
4	90 mm - 180 mm
5	180 mm - 360 mm

coefficient as offered by Federer (1955) is

$$b' = \frac{\Sigma(n_i-1)\log x_i \log V_{x_i} - [\Sigma(n_i-1)\log V_{x_i}\,\Sigma(n_i-1)\log x_i]/\Sigma(n_i-1)}{\Sigma(n_i-1)(\log x_i)^2 - [\Sigma(n_i-1)\log x_i]^2/\Sigma(n_i-1)} \tag{19}$$

which is used as the estimate of the heterogeneity index.

Sampling Study

Once the optimum size sampling unit was found, an estimate of the variance associated with that size was used to determine the sample size to be taken. Given a prescribed interval for a desired level of confidence, the sample size was determined from the half length of the confidence interval as

$$L = ts/\sqrt{n} \tag{20}$$

where s is the square root of the variance and t is the value of Student's t. Setting L equal to the desired half length interval, the required number of quadrats to be sampled was determined.

The entire area under consideration was divided into nine 100m × 100m squares, each called a hectar. The nested sample was taken by randomly selecting a number of hectars and then selecting an equal number of subsamples from each hectar. The results from equation (20), to be discussed later, showed that approximately 100 samples should be taken. It was then decided to choose 25 samples from each of four hectars.

Sampling was done by the sampler walking a random number of paces between successive samples. The number of paces were determined from lists of 25 two digit random numbers drawn for each of the hectars. It was desirable that the sample be representative of the hectars sampled in the north-south direction, as well as the east-west. For this reason each hectar was divided into approximately twelve parallel imaginary strips, each ten paces wide, in the east-west direction. For this purpose a second list of one digit random numbers was selected. The sampler began at the southwestern corner of each hectar and selected a number from the second list and walked that number of paces north. The first sample was then located by determining the number of paces to be taken east from the first list. After each sample was located, the next two digit random number determined the number of paces to the next sample. Upon reaching the eastern boundary of the hectar, he would walk north ten minus the number of paces previously taken north, then choose another random number of paces to be taken north from the second list, and resume sampling by walking west. If the northern boundary was reached before the 25 samples were taken, he would finish his sample by turning back and use the second list to determine his steps to be taken south. Since the expected number of steps to be taken between successive samples was 50, allowing for some error because the sampler might not have walked due east or west, the samples were therefore expected to represent the entire hectar in the north-south direction.

CHAPTER IV

RESULTS AND DISCUSSION

Uniformity Trial

The uniformity trial data from all rock size classes combined, were tabulated for the various quadrat sizes. For each different quadrat, its size in basic units, shape, and number of samples taken from the site, are presented in Tables 2 and 3 for sites one and two respectively. The variance per basic unit for each quadrat, computed as the ratio of its variance to the square of its size, is also presented in Tables 2 and 3. The plots of log of variance per basic unit (V_x) against the log of quadrat size (x) are shown in Figures 1 and 2 for sites one and two respectively. The estimated heterogeneity indices for sites one and two were 0.586 and 0.635 respectively, with 0.035 and 0.046 as their respective standard errors. Tables 4 and 5 show the analysis of variance for each of these two weighted regressions. The respective coefficients of determination for sites one and two were 0.899 and 0.861. Though these two values are highly significant, there is no means of determining their validity as none of the literature presents an analysis of variance for uniformity trial studies.

Table 2. Number of Basic Units, Shape, Variance per Basic Unit and Coefficient of Variation for Rocks from Site 1

No. of Basic Units	Shape in .5m Squares	No. of Samples	V_x	C.V.
1	1 × 1	144	571	.415
2	1 × 2	72	435	.362
2	2 × 1	72	396	.345
3	1 × 3	48	339	.319
3	3 × 1	48	265	.282
4	1 × 4	36	275	.287
4	2 × 2	36	346	.323
4	4 × 1	36	242	.270
6	1 × 6	24	214	.254
6	2 × 3	24	251	.275
6	3 × 2	24	227	.262
6	6 × 1	24	160	.220
8	2 × 4	18	201	.246
8	4 × 2	18	209	.251
9	3 × 3	16	162	.221
12	1 × 12	12	159	.219
12	2 × 6	12	174	.229
12	3 × 4	12	134	.201
12	4 × 3	12	167	.224
12	6 × 2	12	146	.210

Table 2, Continued

No. of Basic Units	Shape in .5m Squares	No. of Samples	V_x	C.V.
12	12 × 1	12	62	.137
16	4 × 4	9	130	.198
18	3 × 6	8	114	.186
18	6 × 3	8	91	.166
24	2 × 12	6	149	.212
24	4 × 6	6	111	.183
24	6 × 4	6	79	.154
24	12 × 2	6	55	.128
36	3 × 12	4	90	.165
36	6 × 6	4	65	.140
36	12 × 3	4	41	.111
48	4 × 12	3	111	.183
48	12 × 4	3	19	.075

Table 3. Number of Basic Units, Shape, Variance per Basic Unit and Coefficient of Variation for Rocks from Site 2

No. of Basic Units	Shape in .5m Squares	No. of Samples	V_x	C.V.
1	1 × 1	144	808	.364
2	1 × 2	72	566	.305
2	2 × 1	72	617	.318
3	1 × 3	48	478	.280
3	3 × 1	48	427	.265
4	1 × 4	36	357	.242
4	2 × 2	36	441	.269
4	4 × 1	36	302	.222
6	1 × 6	24	280	.214
6	2 × 3	24	375	.248
6	3 × 2	24	308	.245
6	6 × 1	24	312	.226
8	2 × 4	18	292	.219
8	4 × 2	18	205	.183
9	3 × 3	16	257	.205
12	1 × 12	12	96	.126
12	2 × 6	12	246	.201
12	3 × 4	12	264	.208
12	4 × 3	12	160	.162
12	6 × 2	12	244	.200
12	12 × 1	12	76	.112

Table 3, Continued

No. of Basic Units	Shape in .5m Squares	No. of Samples	V_x	C.V.
16	4 × 4	9	163	.218
18	3 × 6	8	167	.165
18	6 × 3	8	209	.185
24	2 × 12	6	86	.119
24	4 × 6	6	138	.150
24	6 × 4	6	221	.190
24	12 × 2	6	45	.086
36	3 × 12	4	78	.113
36	6 × 6	4	153	.158
36	12 × 3	4	28	.058
48	4 × 12	3	50	.091
48	12 × 4	3	20	.057

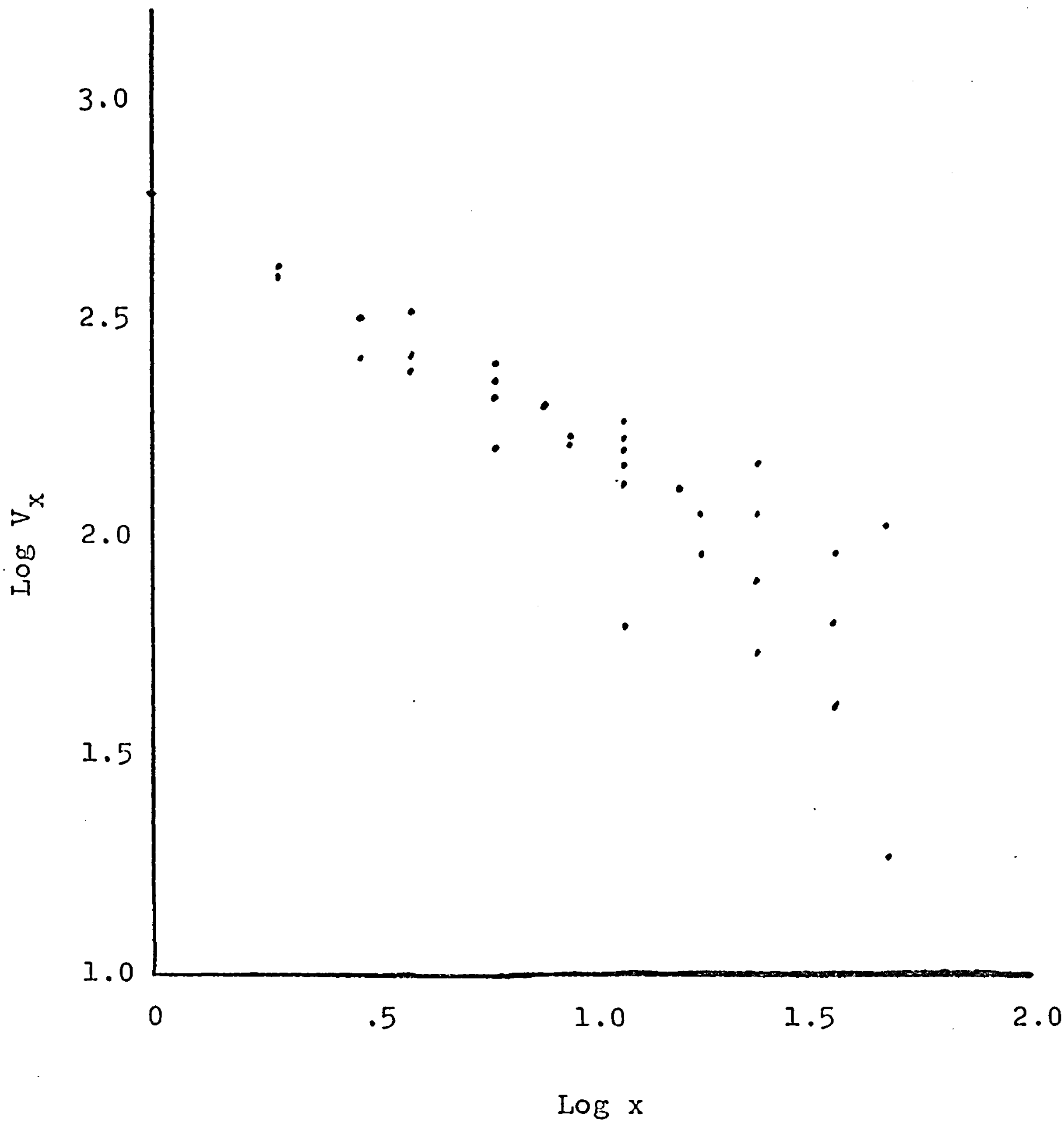

Figure 1. Plots of the log of variance per basic unit against log of quadrat size from site 1 data.

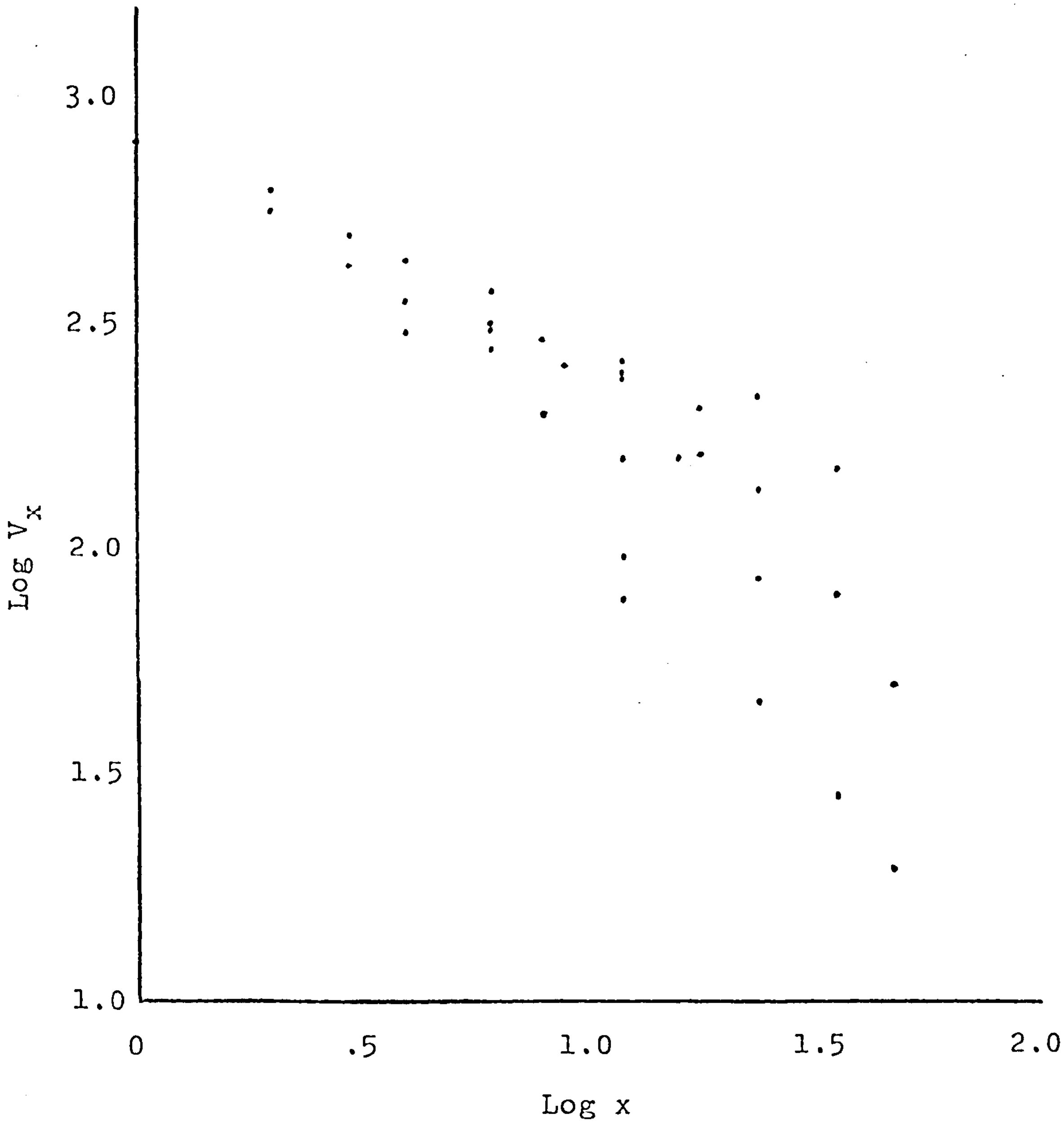

Figure 2. Plots of the log of variance per basic unit against log of quadrat size from site 2 data.

Table 4. Analysis of Variance for Regression of Site 1 Uniformity Trial Data

Source	d.f.	SS	MS
Regression	1	41.673	41.673
Error	31	4.666	.151
Total	32	46.339	

Table 5. Analysis of Variance for Regression of Site 2 Uniformity Trial Data

Source	d.f.	SS	MS
Regression	1	48.852	48.852
Error	31	7.917	.255
Total	32	56.769	

The two regression coefficients did not differ by Student's t at a significance level of 0.05. A significance level of 0.05 was used throughout this paper. The data from sites one and two were pooled into one regression. The analysis of variance for this pooled data, presented in Table 6, yielded an r^2 of 0.819, also highly significant. The pooled heterogeneity index was thus estimated as 0.611 with 0.036 as its standard error.

The cost ratio, K_1/K_2, in equation (4) was found to be 0.67. One minute was the average time to locate a sample, and 1.5 minutes to count all rocks within one basic unit. The optimum quadrat size was then found to be 1.047. A basic unit of one half meter square was therefore used in the sampling study.

Since this uniformity trial was performed with all rocks larger than 11.2mm, rocks of different sizes may give an entirely different estimate of the heterogeneity index. Other studies, such as Kuehl and Kittock (1969) for cotton yields, also show wide dispersions for similar experiments performed at different locations. Trials in other areas may also cause the cost ratio to differ. For these reasons, values for the optimum sample size for a range of b' and the K_1/K_2 proportion are presented in Table 7.

Table 6. Analysis of Variance for Pooled Uniformity Trial Data

Source	d.f.	SS	MS
Regression	1	90.445	90.445
Error	62	19.989	.322
Total	63	110.434	

Table 7. Optimum Quadrat Sizes for a Range of Cost Proportions and Heterogeneity Indices

b′	K_1/K_2				
	.25	.5	.75	1.0	1.5
.3	.107	.214	.321	.421	.622
.4	.167	.333	.500	.667	1.000
.5	.250	.500	.750	1.000	1.500
.6	.375	.750	1.125	1.500	2.250
.7	.583	1.167	1.750	2.333	3.500
.8	1.000	2.000	3.000	4.000	6.000
.9	2.250	4.500	6.750	9.000	13.500

Sampling Study

The mean number of rocks per quadrat were 57.65 and 78.12 for sites one and two, with 571.24 and 808.36 as their respective variances. It was desired to arrive at a 0.95 confidence interval on the true mean number of rocks per quadrat. The half length of this interval was then chosen so as not to exceed 15% of the true mean. Assuming the true mean to be approximated by the two means already found, 15% of the larger mean or five rocks was used as the size of the half length interval.

The variances from the two sites were found to be homogeneous by their F ratio. Their pooled variance, 689.80, was therefore used to estimate the population variance. Utilizing equation (20), the sample size, n, was thus found to be 102.9. A nested sample of 100, 25 quadrats from each of four hectars, was then taken.

The variances of rocks for all size classes within each hectar were found to be homogeneous by the F max test. With this assumption of analysis of variance appearing to be met, an analysis of variance, presented in Table 8, was performed on the four hectars data. The two variance component estimates, $\hat{\sigma}_s^2 = 2692$ for samples and $\hat{\sigma}_h^2 = 220.7$ for hectars, were employed in equation (6), yielding a variance of the mean of 82.10. As was anticipated after sampling the

Table 8. Analysis of Variance of Rocks from All Size Classes Taken from Four Hectars

Source	d.f.	SS	MS	EMS
Hectars	3	24,630	8210	$\sigma_s^2 + 25\sigma_h^2$
Samples within hectars	96	258,432	2692	σ_s^2
Total	99	283,062		

four hectars, there was a great increase in this variance from the estimated variance of the mean of 100 quadrats from sites one and two of the uniformity trial, whose estimates became 5.71 and 8.08 respectively. This sample would give $(\hat{\sigma}_h^2 + \hat{\sigma}_s^2)/100$ as the estimated variance of a sample mean of size 100 using simple random sampling. This variance 29.13, then, is considerably lower than that of the nested sample.

The variance component estimates were next employed in equations (7) and (10) to estimate the optimum allocation for a nested sample from the observed data. The two costs, c_1 and c_2, are the time to travel between hectars and the time to count all rocks within one sampling unit plus the time to walk to the next unit. In this study average times were found to be 10 minutes and 2.5 minutes for c_1 and c_2 respectively. Thus, utilizing the above values, equation (7) gives 7.0 as the optimum number of sampling units per hectar. For a desired half length of a 0.95 confidence interval equal to five rocks, the prescribed variance becomes 6.25 from equation (20). The optimum number of hectars, by equation (10), should then be 97. The large variance among hectars would account for an optimum allocation of hectars exceeding the number of samples to be taken from each hectar. Table 9 presents a range of optimum sampling units per hectar for a range of c_1/c_2 ratios and Table 10 presents various optimum numbers of hectars for a range of n_2 and desired variances.

Table 9. Optimum Number of Samples to be Taken per Hectar for a Range of c_1/c_2 Ratios

c_1/c_2	Number of Samples
2	4.9
4	7.0
6	8.6
8	9.9
10	11.0
12	12.1

Table 10 presents various optimum numbers of hectars for a range of n_2 and desired variances.

Sampling Distributions

Data of rocks for all size classes from each of the uniformity trial sites tended toward unimodal symmetric distributions. Their means were 57.65 and 78.12 for sites one and two, with 23.90 and 28.43 as their respective standard deviations. Using these values as parameters, each of the distributions fit a normal approximation by the Kolmogorov-Smirnov test for goodness of fit.

The F ratio of the hectar mean square to the within hectar mean square showed significance from the analysis of variance of Table 8. It then appeared that among the four hectar means of 71.2, 95.5, 98.5 and 115.2, there were some differences. The distributions of these data from each of the hectars were not tested to fit any distributions, as their histograms did not tend toward any regular shape.

Rocks from the smallest size class were eliminated and the remaining rocks tested for distributional properties. The sample distributions of these rocks from each of the two uniformity trial sites tended toward unimodal symmetric distributions. Table 11 presents these observed frequencies for each of the sites. Their means 5.28 and 9.42 for sites one and two differ by Student's t, as do their variances of

Table 10. Optimum Number of Hectars to be Taken for a Range of Optimum n_2 and Preassigned Variances

n_2 \ V	5	10	15	20
4	178.8	89.4	59.6	44.7
6	133.8	66.9	64.6	33.5
8	111.4	55.7	37.2	27.9
10	98.0	49.0	32.6	24.5
12	89.0	44.5	29.6	22.3

Table 11. Frequency Distributions of Rocks from Size Classes Two and Larger for Each of the Uniformity Trial Sites

Numbers of Rocks	Frequency Site 1	Frequency Site 2
0	1	1
1	10	1
2	12	5
3	15	6
4	26	7
5	23	14
6	20	14
7	10	11
8	9	14
9	7	12
10	2	13
11	3	4
12	2	8
13	1	2
14	1	4
15	0	6
16	1	3
17	0	9
18	1	2
19		2
20		2

Table 11, Continued

Numbers of Rocks	Frequency Site 1	Frequency Site 2
21		1
22		0
23		2
24		0
25		1

9.17 and 23.55 by their F ratio. These means and their standard deviations of 3.03 and 4.85 were used to fit the normal approximation to this data. Again the Kolmogorov-Smirnov goodness of fit accepted these as having come from discrete distributions closely approximated by the normal.

Elimination of the minimum size class of the rocks from the four hectars showed the four groups having homogenous variances by the F max test. An analysis of variance of this data is presented in Table 12. An F ratio less than one showed the four hectar means of 9.20, 12.52, 13.04 and 13.16, to be alike. The data from all four hectars were then combined, giving a sample mean of 11.98 and variance of 124.48. This observed frequency distribution with class ranges of five is presented in Table 13. The obvious skewness implied a test for the Poisson or negative binomial. A quick and easy test for the Poisson is the statistic $(n - 1)s^2/\overline{x}$, which is distributed as a χ^2 with $n - 1$ degrees of freedom. Since the mean and variance of a Poisson are equal and the sample mean is the most efficient unbiased estimate of the common parameter, 11.98 is the best estimate for λ. The value of this test statistic from the observed sample data was 1032; this χ^2 only being exceeded with probability less than 0.0005.

Table 12. Analysis of Variance of Rocks from Size Classes Two and Larger from the Four Hectars

Source	d.f.	SS	MS
Hectars	3	263.40	87.80
Samples within hectars	96	12,060.56	125.63
Total	99	12,323.96	

Table 13. Frequency Distribution of Rocks from Size Classes Two and Larger Combined for All Four Hectars

Number of Rocks	Frequency
0 - 4	33
5 - 9	21
10 - 14	11
15 - 19	14
20 - 24	7
25 - 29	7
30 - 34	2
35 - 39	2
40 - 44	1
45 - 49	1
50 - 54	0
55 - 59	1

The above data was also tested to fit the negative binomial distribution. The chi square goodness of fit test, where $\chi^2 = \Sigma(O - E)^2/E$ is the test statistic, was applied to test for both the Poisson and negative binomial. The values O and E are the observed and expected frequencies. This χ^2 is distributed with d - p - 1 degrees of freedom where d is the number of frequency classes in which data are grouped and p is the number of estimated parameters. The negative binomial, therefore, would produce a χ^2 with d - 3 degrees of freedom, and the Poisson d - 2.

The negative binomial parameter k was estimated as 1.259 by the maximum likelihood method of equation (16). Fitting the negative binomial to this data would have yielded a χ^2 of 74.7 with 56 degrees of freedom (P = 0.048). However, since the expected frequencies were below five in all but the first seven of the 59 classes, the Nass criterion, Pahl (1969), is appropriate for testing goodness of fit. For this, the criterion

$$C = 2 \cdot E(\chi^2) \cdot \chi^2 / Var(\chi^2) \qquad (21)$$

is tested as a χ^2 with q degrees of freedom. The values of $E(\chi^2)$ and $Var(\chi^2)$ are the expected value and variance of the χ^2 given by Pahl (1969) as

$$E(\chi^2) = d - p - 1 \qquad (22)$$

and

$$\mathrm{Var}(\chi^2) = [2(d-1) + (d-1)^2/N + \Sigma(E_i)^{-1}]\left(\frac{d - p - 1}{d - 1}\right) \tag{23}$$

where N is the sample size and E_i is the expected frequency in the i'th class. Pahl (1969) gave the degrees of freedom as

$$q = 2 \cdot [E(\chi^2)]^2(d-p-1)/[(d-1)\mathrm{Var}(\chi^2)] \ . \tag{24}$$

The Nass criterion was 24.71 with 18.52 degrees of freedom. Since this was a nonsignificant χ^2 (P = 0.17), the negative binomial was accepted to fit this data.

The final data to be tested was all rocks larger than 45mm which are size classes three and larger. Data from the uniformity trial had means of 0.549 and 1.201 rocks per quadrat for sites one and two with their variances 0.683 and 2.176. These data were then tested to fit the Poisson and negative binomial. In each case, the sample mean was used as the estimate of λ in the Poisson. The negative binomial parameter k was estimated to be 1.55 by the number of zeroes method for site one data and 1.86 by maximum likelihood for site two. Table 14 shows the observed frequency and expected frequencies for both the Poisson and negative binomial from site one data, and Table 15 presents similar frequencies from site two. Site one data fit the Poisson,

Table 14. Observed Frequency and Expected Frequencies Fitting the Poisson and Negative Binomial to Rocks in Size Classes Three and Larger from Site 1

Numbers of Rocks	Observed Frequency	Poisson Expected	Negative Binomial Expected
0	90	83.38	90.02
1	34	45.65	36.42
2	16	12.53	12.12
3	3	2.59	5.44
4	1		

Table 15. Observed Frequency and Expected Frequencies Fitting the Poisson and Negative Binomial to Rocks in Size Classes Three and Larger from Site 2

Numbers of Rocks	Observed Frequency	Poisson Expected	Negative Binomial Expected
0	57	43.37	57.88
1	40	52.04	40.94
2	31	31.23	22.89
3	6	12.48	11.67
4	4	4.84	5.66
5	3		4.96
6	0		
7	2		
8	1		

yielding a χ^2 of 5.225 with two degrees of freedom (P = 0.08), while site two failed to fit the Poisson, its χ^2 being 15.944 with three degrees of freedom (P < 0.005). Both sites one and two fit the negative binomial distribution. Site one's χ^2 was 1.784 with one degree of freedom (P = 0.20) and 6.368 with three degrees of freedom was the χ^2 for site two (P = 0.10).

An analysis of variance for the larger rocks in size classes three and greater for the four hectar data is presented in Table 16. Again, an F ratio less than one permitted the data to be pooled, as if drawn from a common population. The mean and variance for these rocks were 1.18 and 3.543. The Poisson and the negative binomial were fit to this data. The negative binomial parameter k was estimated as 0.520 by the number of zeroes method. Table 17 presents the observed frequency and the expected frequencies from fitting the Poisson and negative binomial to this data. The Poisson did not fit the data, yielding a χ^2 of 47.5 with three degrees of freedom (P < 0.005), while the negative binomial did fit the data quite well, its χ^2 being 0.950 with three degrees of freedom (P = 0.82).

An index of clumping is only of interest in contagious distributions. Therefore, a discussion of this is limited to rocks larger than 45 mm. The three indices from equation (17) were 0.244, 0.811 and 2.00, for site one,

Table 16. Analysis of Variance for Rocks in Size Classes Three and Larger from the Four Hectars

Source	d.f.	SS	MS
Hectars	3	8.120	2.707
Samples within hectars	96	342.640	3.569
Total	99	350.760	

Table 17. Observed Frequency and Expected Frequencies Fitting the Poisson and Negative Binomial to the Combined Data of Rocks Larger than 45mm from the Four Hectars

<table>
<tr><th>Numbers of Rocks</th><th>Observed Frequency</th><th>Poisson Expected</th><th>Negative Binomial Expected</th></tr>
<tr><td>0</td><td>54</td><td>30.73</td><td>53.94</td></tr>
<tr><td>1</td><td>19</td><td>36.26</td><td>19.46</td></tr>
<tr><td>2</td><td>9</td><td>21.39</td><td>10.27</td></tr>
<tr><td>3</td><td>8</td><td>8.41</td><td>5.99</td></tr>
<tr><td>4</td><td>4</td><td rowspan="7">3.21</td><td>3.65</td></tr>
<tr><td>5</td><td>2</td><td rowspan="6">6.69</td></tr>
<tr><td>6</td><td>1</td></tr>
<tr><td>7</td><td>1</td></tr>
<tr><td>8</td><td>1</td></tr>
<tr><td>9</td><td>0</td></tr>
<tr><td>10</td><td>1</td></tr>
</table>

site two and the four hectars respectively. As these values have not been previously estimated from rock data, they possess little more than comparative value. The indices from sites one and two did not differ by equation (18), but the index from the four hectars was significantly greater than that from each of the two uniformity trial sites. This showed that, as observed, the two uniformity trial sites were quite similar, while sampling over the larger area produced much greater clumping.

The final measure was an estimate of the proportion of ground covered by rocks 11.2mm and larger. Each size class was assigned a class mark, which was the antilog of the mean of the logarithms of the class end points. This value for each size class was then assumed to be the mean size of all the rocks in that class. As an average, all the rocks in a given size class were assumed to be square with each side equal to its class mark. The area covered by a square, having the class mark as the length of its sides, then became the class mark of a frequency distribution of areas. Table 18 presents these values with the mean frequencies of rocks per quadrat in each size class as recorded from the four hectars. The mean ground coverage per quadrat was then estimated as the sum of the products of the frequencies times their respective mean areas for all the

Table 18. Mean Frequencies and Their Associated Variances for the Mean Size and Mean Coverage Observed in Each Size Class

Class	Class Mark	Area Covered	Mean Frequencies	Variance of the Mean Frequencies
1	15.78mm	$249mm^2$	83.12	20.66
2	31.75mm	$1008mm^2$	10.80	.984
3	63.64mm	$4050mm^2$	1.07	.031
4	127.30mm	$16205mm^2$	.11	.0013

size classes. This mean coverage was found to be 37,700mm^2, or 15.1% of the half meter square quadrat.

The above mean ground coverage was estimated from the linear function $\Sigma a_i f_i$, where a_i is the mean area for the i'th class and f_i its mean frequency. Each of the areas are constants, and each of the mean frequencies have variances also presented in Table 18. The variance of the estimated mean ground coverage per quadrat is then

$$V = \Sigma a_i^2 S_i^2 \qquad (24)$$

where S_i^2 is the variance of the mean frequency in the i'th class. This variance was thus found to be 3,120,000. The variance of the percent ground coverage then became 0.00005, since the 15.1% was estimated as the ratio of the mean coverage to the 250,000mm^2 contained in a half meter square. The 0.95 confidence interval estimate for the true percent ground coverage is therefore 13.7 to 16.5%.

The area sampled appeared to be divided into a number of distinct strata. This has been shown by the clumping observed by the sampling distributions of rocks 45mm and larger. The strata tend to vary from those covered solely by small rocks to those densely covered with much larger rocks. Unfortunately, these strata have not been defined as yet. Identification of particular strata and dividing a desert region into percent coverage by each strata would be a possible goal of any future desert pavement coverage study. However, this would be a descriptive approach.

A stratified sample would be the best sampling scheme if the strata could be identified and quantified. Assuming this to be unfeasible, a simple random sample appears to be better than the nested sample, which is no more than an artificial stratification of the area into hectars. The variance of the mean number of rocks per quadrat was 82.10 for the nested sample employed, while the same data, taken as a simple random sample, would have 29.13 as the estimated variance of its mean. The variance among hectar means was therefore too great to allow a nested sample to be used.

This study began with emphasis on all rocks 11.2mm and larger; however, the distribution of rocks with a larger minimum size class seem to possess more interesting properties. The error in measurement due to the inability to distinguish between rocks larger than 11.2mm from those smaller could well have introduced a large error in the sample data. For these reasons, perhaps a future study would consider only rocks of larger size classes.

The uniformity trial, as performed for this study, may well have been misleading, due to the smaller variance observed among adjacent quadrats over a uniform site. Such a uniformity trial with larger rocks would require a greater basic unit size. Yet this could prove prohibitive in a typical Southern Arizona desert, because large areas free of vegetation are extremely difficult to locate.

CHAPTER V

SUMMARY

It was found that a one half meter square quadrat is the optimum size sampling unit in which to sample desert rocks. The method of estimating this quadrat size--a uniformity trial--appears to be invalid for rocks, however. This is due to the variance among adjacent quadrats, as used in a uniformity trial, being considerably less than that among quadrats placed randomly over a large desert region.

Of the area studied, half meter square quadrats had a mean of 95.1 rocks. This was divided into 83.12 rocks between 11.2 and 22.4mm, 10.80 between 22.4 and 45mm and 1.18 greater than 45mm. The proportion of the ground covered by these rocks approximated 15%. Much of the remaining ground, though, was covered by smaller rocks. Consequently, decreasing the minimum size class should greatly increase the percent ground coverage.

A large variability was observed because the ground coverage appeared to be highly stratified as to degrees of density. The definite clumping observed from the sampling distribution of rocks larger than 45mm showed this.

The optimum approach to sampling desert rocks appears to be a stratified sample. This entails identifying and quantifying the various strata, however. A nested or cluster sample, whereby the area is artificially subdivided into unnatural areas of equal size, is an undesirable method of sampling rocks because of the large observed variance among the subdivisions employed in this study. Therefore, a simple random sample appears to be the optimum approach to sampling desert rocks.

REFERENCES CITED

Bancroft, T. A. and T. A. Brindley. Methods for estimation of size of corn borer populations. Pro. X International Congress of Entomology. 2:1003-14. 1956.

Bliss, C. I. and R. A. Fisher. Fitting the negative binomial distribution to biological data. Biometrics. 9:176-200. 1953.

Brim, C. A. and D. D. Mason. Estimates of optimum plot size for soybean yield trials. Agron. J. 51:331-34. 1959.

Cochran, W. G. Sampling techniques. New York, John Wiley and Sons, Inc. 1963.

David, F. N. and P. G. Moore. Notes on contagious distributions in plant populations. Ann. Bot. Lond. N. S. 18:47-53. 1954.

Federer, W. T. Experimental design. The Macmillan Co., New York. 1955.

Greig-Smith, P. Quantitative plant ecology. London. Butterworths. 1964.

Kuehl, R. O. and D. L. Kittock. Estimates of optimum plot size for cotton yield trials. Agron. J. 61:584-86. 1969.

Lane, E. W. Report of the subcommittee on sediment terminology. 1957.

Marcuse, S. Optimum allocation and variance components in nested sampling with an application to chemical analysis. Biometrics. 5:189-206. 1949.

Neyman, J. On a new class of "contagious" distributions, applicable in entomology and bacteriology. Annals of Math. Stat. 10:35-57. 1939.

Pahl, P. J. On testing for goodness-of-fit of the negative binomial distribution when expectations are small. Biometrics. 25:143-51. 1969.

Pielou, E. C. An introduction to mathematical ecology. New York, Wiley-Interscience. 1969.

Smith, H. F. An empirical law describing heterogeneity in the yields of agricultural crops. Jour. Agr. Sci. 28:1-23. 1938.