Random and Non-random Sampling of the Same Site

BRIAN SPURLING

Introduction

The need for probabilistic sampling strategies in archaeological research has been formally recognized since Vescelius (1960) proselytized the use of explicit sampling techniques. Binford (1964) and Rootenberg (1964) offered research designs deploying probabilistic sampling methods at the regional and site levels respectively four years later. By 1967 Ragir (1967) contributed a synopsis of sampling tactics largely influenced by Krumbein's (1965) work in paleontological sampling. Although all of these researchers proffer the rationale supportive of probabilistic sampling in archaeological research, none provide tests of efficiency between judgemental or intuitive sampling methods and probabilistically derived samples, or between alternative probabilistic sampling strategies conducted against a real population of archaeological data. This situation was rectified by Plog (1968), Matson (1970), Thomas (1969, 1973b), Judge, Ebert, and Hitchcock (1973), Matson and Lipe (1973), and Meuller (1974). These studies were generally concerned with statistically appraising the efficiency of differential sampling techniques employed against a universe of known site locations at the regional scale. Published accounts of such comparisons conducted through inter or intra-site excavational programs are scant. Only one case study, unfortunately of a parochial nature is known to the author, that of Michael Blake's excavation of a housepit at EeRk 9 and his subsequent comparison of the accuracy and precision of three probabilistic techniques, various sample sizes, and stratification designs against the recovered artifact population (Blake 1974). It is the intent of this paper to present the results of a similar investigation undertaken at a small coastal shell midden (DcRu 2) located near Victoria, British Columbia during the 1973 field season.

Study Area and Methods

DcRu 2 is situated near the southern entrance to Esquimalt harbour south and east of Victoria, British Columbia at the northern margin of Esquimalt Lagoon. It is separated from the terminus of Coburg Peninsula by a shallow tidal channel which drains and replenishes the lagoonal waters. The site is colloquially known as the Esquimalt Lagoon site for obvious reasons (Figs. 23, 24).

The surface area of DcRu 2 is approximately 4228 m^2. of which 3412 m^2. is amenable to sampling (a road runs through and covers a part of the southern section of the site). The site represents 6432 ± 130 m^3. in total volume of which 5118 ± 104 m^3. is accessible for excavation (80% limits).

DcRu 2 is within the historic-ethnographic tenure of the Straits Salish Songhee and, at present, under the tenure of the Historic Sites Branch of the Government of Canada.

The excavational history of the site spans the field seasons of 1972 and 1973. During the months May to September of 1972 Mr. Ernest K. Oliver directed five students in excavations at the site. This program was recommenced in October of 1972 for the purposes of a six-week field school course in

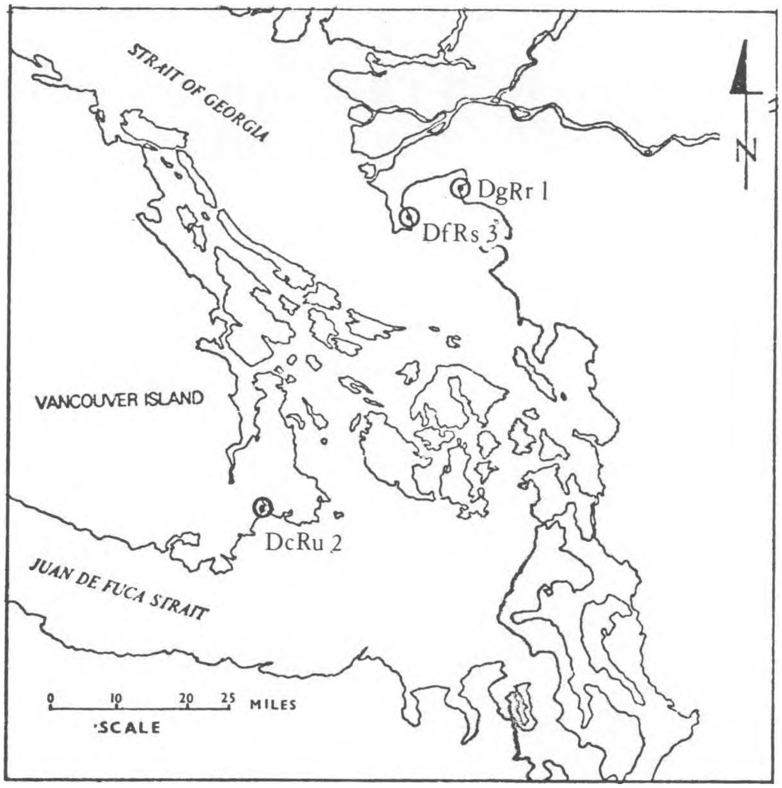


Fig. 23. *Location of the Esquimalt Lagoon site, DcRu 2. The locations of the Whalen site, DfRs 3, and the Crescent Beach site, DgRr 1, are also shown.*

archaeological methods offered by the University of Victoria again under the direction of Mr. Oliver. The author directed a crew of from five to seven members in further excavation during the months of May to August 1973.

The cumulative retrieved technological collection is grossly assignable, with some reservation, to the *San Juan phase* (Carlson 1970) or Mitchell's (1971b) *Montague Harbour III* taxon. "Diagnostic" of the *San Juan phase* are unspecified quantities of herring rake barbs, unilaterally barbed bone points, small unbarbed bone points, sandstone abraders, ground nephrite adze blades, valves, composite socketed harpoon heads, thin ground slate knives, thin triangular ground slate points, and antler wedges. Artifact classes occurring in quantity are composite fish hook barbs and items fashioned from split elk and deer long bones. Other artifact types present but in low frequencies are chipped stone items (Carlson 1970). Similarities are also adduceable to King's (1950) *Late Phase* at Cattle Point, Bryan's (1963) *Late period* components from northern Puget Sound, or Kidd's (1969) Late Component. The collection from DcRu2 cannot be assigned on a wholesale basis to the above taxon without the following qualifications. The presence of one specimen of a type assignable to an earlier regional taxon and the percentage that one class of cultural items represents in relation to the retrieved 1973 collection argue for 1) the presence of two components at the site and/or 2) recognition of an earlier date for the inception of the *San Juan phase*.

The *San Juan phase* here is used as a "catchall" temporal/spatial/cultural taxon inclusive of Mitchell's, Kidd's, and Bryan's complexes. The appropriateness of this lumping and, indeed, of the use of the concept of phase as applied to archaeological assemblages of the area in general will go unquestioned. It is beyond the scope of the present paper to undertake a justification of the above assignation of the use of the phase concept. With regard to the anomalous items present in the DcRu 2 collection of 1973 the proximal section of a bone harpoon fragment with bilateral line guards exhibiting posterior and anterior incisions

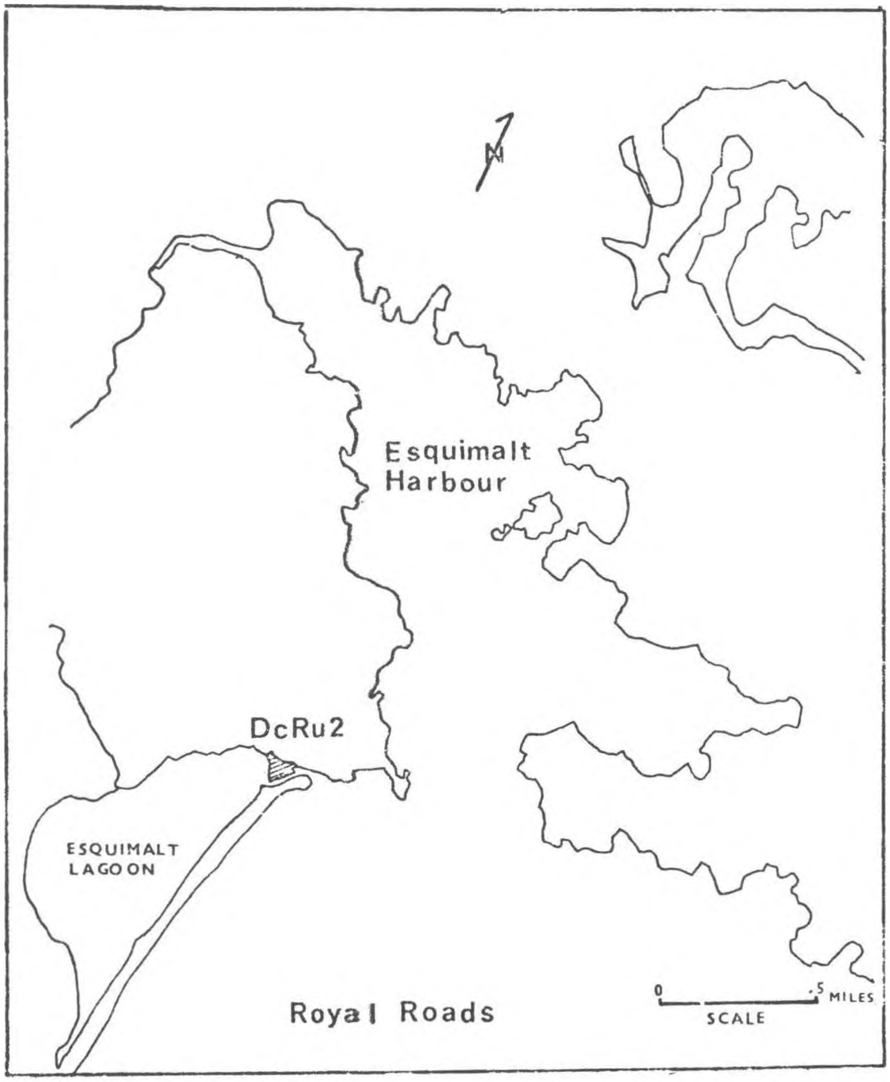


Fig. 24. *DcRu 2 at the northeast end of Esquimalt Lagoon.*

was recovered (McMurdo's Class i, Type I, sub type c 1972). Similar harpoons occurred at the Garrison site on San Juan Island and were assigned to the Marpole phase (Carlson 1960: 586–89). Items representative of a chipped stone industry, i.e., 2 cores, 3 points, 2 scrapers, and 2 flakes, constitute 6.1% of the 1973 collection, a frequency not comfortably consonant with the definition of the *San Juan phase.*

Chronologically the *San Juan phase* "dates from at least A.D. 1200 to the time of European contact . . . (and thus) . . . represents the protohistoric culture of the Straits Salish (Carlson 1970: 120). The *Marpole phase* (400 B.C. to 400 A.D.) precedes this. Thus, superficially, an 800 year lacunae exists between these two phases. However excavations at Fossil Bay undertaken by Kidd (1964) and at Dionisio Point by Mitchell (1971b) have produced San Juan type components dating to 436 A.D. and 550 A.D. respectively. The acceptance of these dates, although controversial at present, would indicate an earlier appearance than hitherto recognized of a culture comparable to that of the region's historic inhabitants. Thus the typologically early technic items of the 1973 DcRu 2 collection may suggest the assignment of the earlier depositional events at the Esquimalt Lagoon site to the above grey period in the chronology of the area. Whether two components are represented by the collection or whether the beginning of the *San Juan phase* should be temporally moved back on the basis of the recovered artifact sample of DcRu 2 are questions which must await the results of radiometric assay.

Having approximately positioned DcRu 2 in time and space we may now turn to the research design undertaken in 1973. A number of hypotheses were tested against the cultural and ecological residue present at the site. The one under concern here is the sufficiency of the sampling strategies deployed against such sites. As with other experiments in the social sciences, excavations of archaeological deposits are non-replicable. The act of recovery is simultaneously an act of destruction. The optimal sampling technique, in terms of retrieval, is thus complete excavation. This can rarely, if ever, be undertaken due to exigencies of time, funding, available trained personnel, and the desirability of leaving parts of sites unexcavated for future research. In lieu of complete sampling it was determined, in 1973, that since the site had been excavated the previous year the deployment of a sampling technique different than that utilized in 1972 would yield data which could be used to compare sampling strategies.

The senior investigator of the previous year had employed a judgemental sampling design based on a trench system. The design was judgemental in that no probabilistic criteria were employed in making the decision as to where excavation units were placed. The stated rationale behind the sampling strategy employed was that a salvage situation existed and that "because the east face of the midden was the most endangered by wave erosion, it was decided to concentrate the sampling in this area" (Oliver 1973). Other criteria likely entering into the choice of excavated area were, from my own observations, an absence of large vegetation and the presence of a large open area running on a north-south axis which would facilitate the taking of provenience data. However the ultimate factors behind judgemental sampling are psychological and, for our purposes, unknowable. Eleven 1 X 2 m. and three 1 X 3 m. excavation units were opened in this area near the eastern face of the midden. These units were staggered along and perpendicular to a north-south axis.

The strategy of sampling decided upon in the 1973 excavation program was an element sampling technique known as simple random sampling (Cochran 1963; Meuller 1974). This scheme was chosen in order to use statistical tests which assume a normal distribution, to reduce the complexity accruing to cluster samples due to the fact that ratio estimates are necessary in the computation of sample statistics, and because no site surface characteristics were available to stratify the site. In element sampling the quadrats are regarded as the elements and the cultural item frequencies, attributes of the elements which are amenable to the use of statistical tests dependent upon the assumption of normality (c.f. Thomas 1973b: 12–13).

In order to simple random sample (SRS) the site, its perimeter was first determined, the site was then mapped and gridded into 2 X 2 m. frames which were numbered. A table of random numbers was then consulted and the units drawn were excavated consecutively throughout the summer. The frame size was dictated by both safety and statistical considerations. In sampling ecological communities the rule of thumb for the selection of frame or quadrat size is to choose the smallest possible, "relative to the type of vegetation and to the practicability of the enumeration of such a quadrat size" (Kershaw 1973:32). Although the enumeration of cultural items should not prove, in a study of this nature, a constraint on size of the sampling frame, the encounter of problems analagous to the effect of vegetational patterning in ecological populations is common in archaeological sampling. As to what comprises a natural sampling unit or clump of cultural items in a shell midden has yet to be determined. While not denying that mosaic patterning exists in such archaeological deposits, we chose to ignore this possibility and opt for that frame size which represented the minimal dimensions in which crew members could work with the highest margin of safety. Independent 2 X 2 m. excavation units were selected as pits of this size when sunk into unconsolidated midden matrix that may achieve depths of up to 3 m. or more represent the minimal size that crew members work in with both comfort and security.

A 0.76% sample of the entire site by volume was achieved by SRS. If that portion of the site inaccessible to sampling is deleted the sample size is increased to 0.93% by volume. Thankfully none of the randomly selected units coincided with the roadway or shoulder.

The difference between the judgemental sample design's dispersion and that of the simple random sample design is illustrated in Figure 25. The computation of R values by Clark and Evans (1954) nearest-neighbour analysis produced a coefficient of .23 for the judgemental sample and 1.33 for the SRS. Dis-

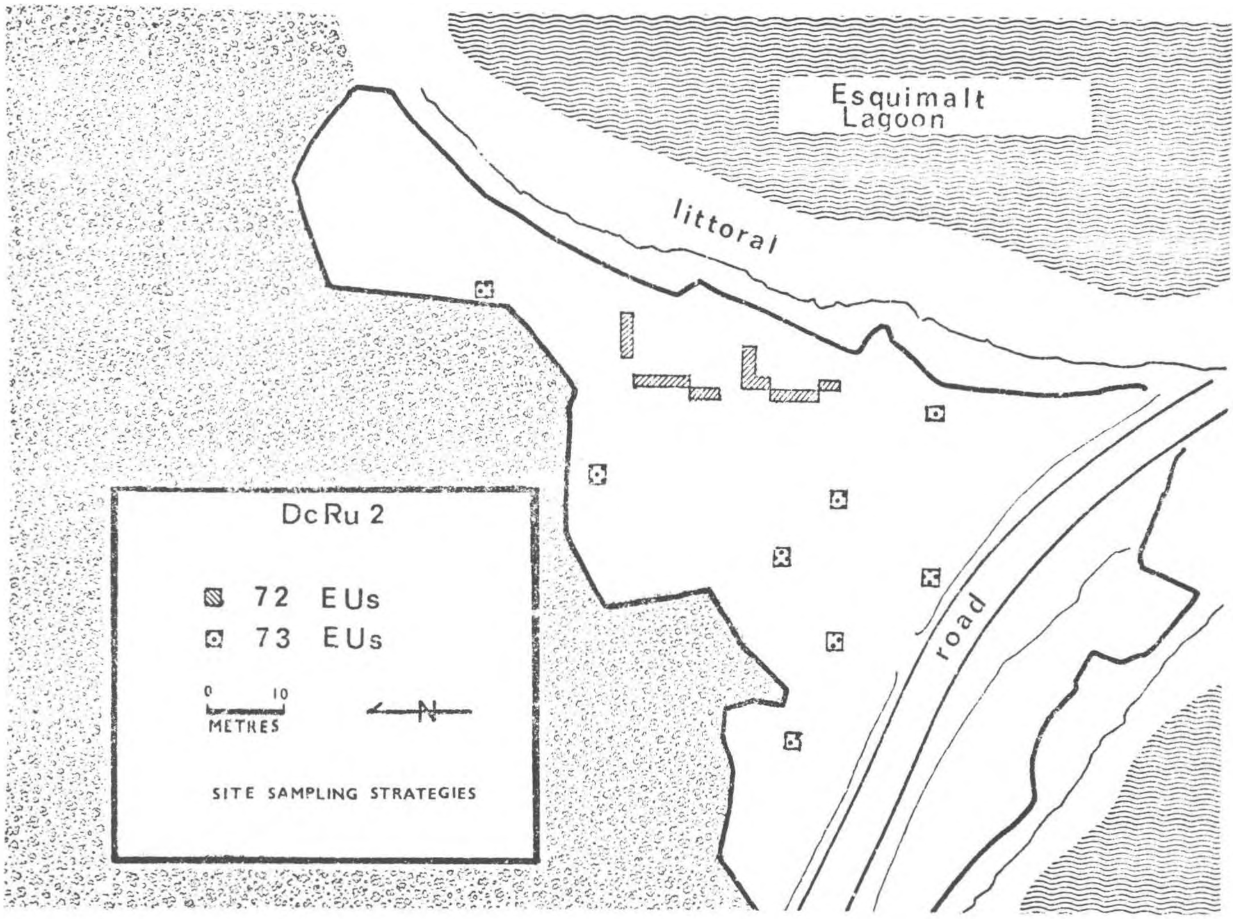


Fig. 25. *Site map of DcRu 2 showing random and non-random excavations.*

tance measurements were made from the centre of each excavation unit to its nearest neighbour. Thus what is intuitively apprehensible from the scrutiny of Fig. 25, i.e., that the judgemental sample is of a clustered pattern and the SRS is dispersed randomly with regard to quadrat placement on the horizontal dimension, is confirmed by nearest neighbour analysis.

As the research design of the 1973 investigation called for comparibility between the sample recovered judgementally the year before and the probabilistically obtained SRS care was taken that the two samples be equivalent in volume. The 1972 excavation program removed approximately 35.3 m^3. of matrix. The program undertaken in 1973 resulted in the removal and screening of approximately 48.0 m^3. of deposition. A Model II single classification anova with unqual sample sizes indicates a significant ($P < 0.05$) added variance component among derived samples for the volume of excavated units of each sampling strategy (Sokal and Rohlf 1969:208–9). This is obviously attributable to the difference in quadrat sizes between the sampling strategies deployed. When the 2 X 2 m. units were divided into two 1 X 2 m. excavation units and volumes obtained again compared an *F* value of 2.32 was computed. As $F_s < F_{.05(1.29)}$ ($F_{.05(1.29)}=4.18$) the null hypothesis is accepted, i.e., the means of the two series are not significantly different. This threshold is considered significant enough for comparing the derived samples.

A two-sample case of a one-way anova (Yeates 1974), a linear regression and two Q mode principal components analyses were used to compare the technic samples recovered by the two sampling strategies. The OTU's selected from between sample comparison are artifact type frequencies because, at the present stage of analysis, they are the most tractable data set. The two inferential statistical methods and multi-variate technique mentioned above were utilized against this data. The precision of the estimates can be compared, if desired, by the use of the within-group MS of the table presented below (Cochran 1963:15; Sokal & Rohlf 1969:195–7).

The threshold data utilized in the 3 analyses described below are presented in Table 1.

One-way analysis of variance compares two different estimates of variation which cumulatively can be employed to calculate the variance of the presumed normally distributed population from which the samples were drawn. The null hypothesis is that the two samples were drawn from the same population with similar variance estimates for each sample which are also no different than the variance estimates for the population. The research hypothesis is that both the judgemental and probabilistic samples differ significantly (Yeates 1974:132–35). The accompanying anova table contains the results of this analysis.

Anova Table

Source of variation	df	SS	MS	Fs
$\bar{Y} - \bar{\bar{Y}}$ Between groups	1	2.65	2.65	.0863
$Y - \bar{Y}$ Within groups	72	2210.	30.69	
$Y - \bar{\bar{Y}}$ Total	73	2212.65		
				$F_{.05(1,72)} = 4.0 > n > 3.92$

Since $Fs << F_{.05(1,72)}$, the null hypothesis is accepted. The means of the two series are not significantly different; that is the two samples do not differ in their technic composition (Sokal and Rohlf 1969:218–219).

The data in the Table were also subjected to simple correlation and regression. The derived Pearsonian product-moment correlation coefficient is $r = .84$ which, when squared, gives a coefficient of determination of 71% which allows the conclusion that 71% of the variation in the judgementally derived sample is associated with the variation in the probabilistically derived sample. Even though the sample size is quite small a test of the standard error of estimate of the regression coefficient produced a value of 8.90($t_{.99(35)} = 2.46 > n > 2.33$). Thus the assumption of normality seems to hold (Yeates 1974:78–9). The best-fit straight-line relation between the two sampling strategies is presented in figure 26.

Principal components analysis is concerned with describing the underlying structure manifested by a group of variables with the assumption that all the variation in a given population is contained within the variables defining the population, i.e., the total matrix variance equals unity. In the analysis carried out against the DcRu 2 sampling programs two Q-mode analyses were employed utilizing the BMDO2M Regression on Principal Components program.

In the first analysis the two sampling strategies were treated as variables and compared across 37 cases comprised of the frequencies of technic items as presented in Table 1. The 2 eigenvalues or sum of the squares of correlation coefficients between each variable and the resolution vector accounting for the largest proportion of total variance were 1.84 and .1593 accounting for 92% and 8% of the total variance respectively. The eigenvector matrix presented below indicates that the variation contributed by each sample is equal and minimal:

Eiqenvectors		**1**	**1**
judgemental sample	1	−0.7071	−0.7071
cluster sample	2	−0.7071	0.7071

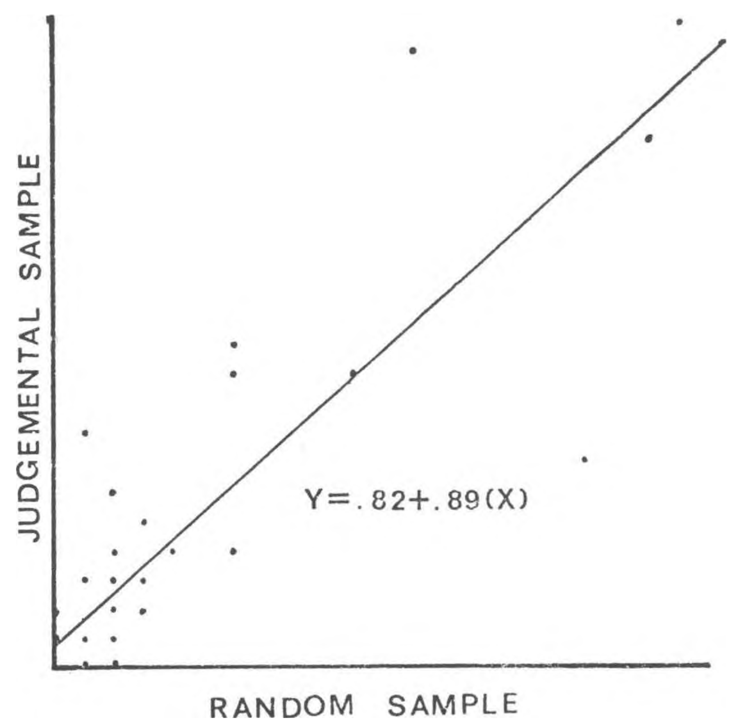
For the second principal components analysis the technic item frequencies were employed as variables and the sampling strategies as cases. In order to have the data conform to the limitations of the program the 37 technic item classes were reduced to 20 variables and the number of cases expanded by 3 by pooling the two samples to form the 3rd case. The new data matrix is presented in Table 2.

Two components were extracted with eigenvalues of 16.0211 and 3.9788 (where matrix variance = 20 = number of variables as matrix diagonals are 1.0's) accounting for 80% and 20% of the total variance respectively. A cumulative frequency graph of the eigenvalues for variables on the first two components are given in figure 27. No group of variables contributed more than 6% to the variance of the first component. This is a vector which apparently reflects homogeneity of samples. It is of interest however that 33% of the remaining 20% of population variance contained in the second component is accounted for by chipped stone items which, as has been suggested earlier, may argue for an earlier chronological placement of the early depositional events at the site. I conclude from this analysis that little variation in technic item recovery between the two sampling strategies exists.

TABLE 1
Data, Means, and Standard Deviations for a Two-Sample Case of a One-Way Analysis of Variance

Type	Threshold value	
	Judge	SRS
Dentalium	1	0
Worked Shell	6	2
Ochre	1	0
Incised Bird Bone	2	2
Ulna Awl	2	1
Long Bone Awl	1	3
Bird Bone Point	10	6
Bird Bone Bi-point	4	4
Bird Bone Fragment	4	2
Mammal Bone Bi-point	21	12
Mammal Bone Point (short)	22	21
Mammal Bone Point (long)	4	6
Curved Bone Object	1	1
Mammal Bone Chisel	1	2
Mammal Bone Fragment	18	20
Valve	8	1
Unilaterally Barbed Point	3	3
Bone Harpoon	0	1
Decorated Object	0	1
Antler Plug	2	0
Antler Point	1	1
Antler Wedge	10	10
Antler Tine	4	2
Antler Fragment	11	6
Core	0	2
Stone Point	2	3
Side Scraper	0	2
Flake	1	2
Slate Blade	2	3
Slate Point	3	1
Slate Fragment	5	3
Hammerstone/Grinder	0	1
Celt	1	1
Shist	1	1
Handmaul	1	1
Abrader	7	18
Mean	4.32	3.94
Standard Deviation	5.68	5.39

Fig. 26. The best-fit straight line relation between the two sampling strategies.



Discussion

Throughout this analysis the research hypothesis has been that differences between the sample obtained by judgemental and probability sampling strategies, other than sums and means, should exist and be detectable by statistical procedures. That the tests deployed convey that the samples are drawn from the same population is not intuitively offensive; however the insignificance between variations of the sampling strategies is disturbing theoretically. The implication of the null hypothesis is, at this admittedly preliminary stage of analysis, that, as far as the recovery of technic items from DcRu 2 is concerned, one would be served equally well by judgement sampling the Esquimalt Lagoon site as by using SRS.

There are obvious advantages to random sampling. Deployment of parametric statistics in inter- and intra-site comparisons is permitted by, indeed is contingent upon, the use of randomly drawn samples. Probability samples can be exploded to generate estimated parameters, within confidence limits, of site volume, artifact type densities, etc. Demographic determinations are similarly facilitated through probabilistically recovered samples of faunal remains (Shawcross 1972).

But there is a trade off. In terms of the present study the 1973 investigation lost some stratigraphic control as compared to that gained in the 1972 program. When the excavation of a shell midden with internally complex stratigraphy is undertaken this loss may prove considerable. The use of a trench

TABLE 2
Modified Data Matrix for Principal Components Analysis

Variable	Case		
	element	SRS	pooled
1. Shell	2	7	9
2. Decorated object	3	3	6
3. Antler	20	28	48
4. Awl	4	3	7
5. Bird bone point	10	14	24
6. Bird bone fragment	2	4	6
7. Mammal bone point	39	47	86
8. Mammal bone fragment	20	18	38
9. Curved bone object	1	1	2
10. Bone chisel	2	1	3
11. Barbed point	3	3	6
12. Toggling harpoon component	1	8	9
13. Ground schist	1	1	2
14. Ground stone (misc.)	2	1	3
15. Flake	5	1	6
16. Chipped stone point	3	2	5
17. Core	2	0	2
18. Ground slate	7	10	17
19. Celt	1	1	2
20. Abrader	18	7	25
Total	146	160	306

system does provide vertical control over a larger area than simple random sampling.

Both the 1972 and 1973 sampling strategies however are at a deficit when questions as to the horizontal patterning of cultural items in a shell midden such as DcRu 2 are addressed. As adumbrated above it is, at present, unclear as to what represents a natural sampling unit of middens of the ilk of DcRu 2. Logistically and ethnographically it is unlikely that DcRu 2 was occupied for a sufficient portion of the seasonal round of the Songhee to warrant the construction of permanent architecture. Determinations of social distances, activity areas, and other factors influencing the deposition of cultural items are lacking. Thus we are left with attempts at detecting redundant associations of cultural items. Although a routine for the recognition of such "toolkits" is under construction by the author, the results of this analysis are of too preliminary a nature to be presented. Even should such redundant associations be discovered their emic significance will be doubtful as their recovery is based upon a sampling frame of entirely arbitrary dimensions. Short of large-scale excavation with increased sample size of probabilistically selected areas of such sites as DcRu 2 the author has no immediate solution to this strategic dilemma.

Returning to the comparison at hand we must ask and answer the question as to why two totally

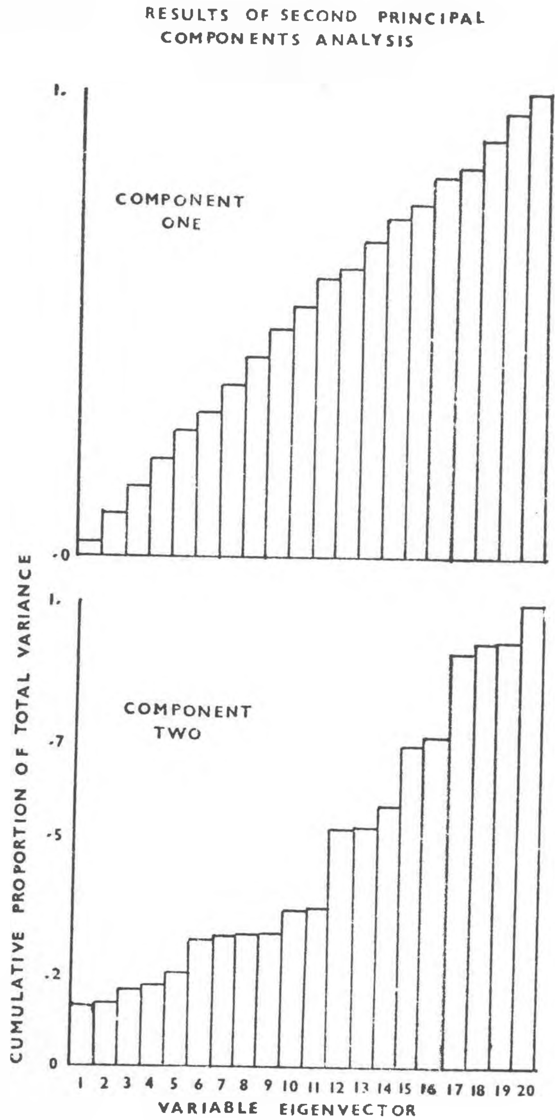


Fig. 27. *Cumulative frequency graph of eigenvector variance of the first two components of the second PCA.*

different sampling strategies should produce technic samples that are essentially similar. Two possibilities seem substantive. Given that the frequency distributions of the OTU's dealt with in this study can be described by a probability generating function (c.f. Neyman 1939, Warren 1971) the two sampling techniques may be encountering a spatially autocorrelated phenomena and essentially sampling the same positions on a cycle time. That is if there are periodicities in the data collected the sampling frames may be encountering the same positions on the oscillation of the dispersion pattern of the data.

A second and related possibility is that the dispersion of the technic items is between random and uniform. Systematic sampling procedures such as by transects (here trenching will be regarded as a special case of transect sampling) produce representative population estimates of very randomly distributed population; random sampling is representative when deployed against a normally distributed population. The technic item patterning of the simple random sample was tested, in a crude fashion, against a simple Poisson distribution. For a Poisson distribution the variance and the mean are equal. Contagious discrete distributions are defined by variances which exceed the mean. When the number of technic items per 10 cm. X 2 m. X 2 m. artifical excavation level are considered for the probabilistically recovered sample a variance: mean ratio of 2.74 is produced. When the index of dispersion (Pielou 1969:91) is calculated and compared to the X^2 distribution it is clear that the judgemental sample is contagiously patterned ($P < .005$). Unfortunately two drawbacks to the use of the variance: mean ratio do not allow me to adduce, with confidence, the empirical nature of the dispersal of the SRS. Jones (1955-56) has suggested that the interpretation of the ratio is unreliable should the mean density of individual items be extremely high or very low. Skellam (1952) has stated that the ratio is too dependent upon the size of the sampling frame or quadrat, a consideration discussed above. A resolution of the question as to the effect of the data distribution of the DcRu 2 technic item population upon the selection of an optimal sampling strategy must await further investigation employing distance-order statistics and attempts to fit the DcRu 2 data dispersion to a series of generalized distributions (Pielou 1969:83–86; Kershaw 1973:135–36).

In conclusion the proximate results of the comparison between judgemental and element sampling of DcRu 2 suggest that, on the basis of the respective technic item frequencies recovered, the probabilistic technique displayed no obivous increase in representativeness over that recovered by the judgemental implicit strategy. Ultimately however, the probabilistic sample is the only collection for which representativeness of the population of cultural items at the site can be claimed. This contention is substantiated on theoretical grounds as the probabilistic sample was obtained, by definition, with the allowance that every technic item of the population had an equal and known probability of selection and recovery. From the probabilistic sample a series of estimates, including those concerned with demography, can be derived. However the element sample derived during the 1973 excavation program at DcRu 2 is as insufficient as the judgemental sample in providing data sets upon which hypotheses can be tested relating to the sociological factors responsible for the horizontal dispersal of cultural items. Until we investigate the theoretical distribution and empirical patterning of our data sets over a wide range of site types recommendations of optimal sampling strategies must necessarily be deferred.

ACKNOWLEDGEMENT

The field research was supported by grant 51340227 from the Opportunities for Youth program. I wish to thank the members of the 1973 excavation program, and J. Nance for assistance in the principal components interpretation, T.K. Peucker for comments concerning sampling in geographical studies, and Catherine Carlson.