AMERICAN ANTIQUITY

VOL. XVIII APRIL, 1953 No. 4

STATISTICAL TECHNIQUES FOR THE DISCOVERY OF ARTIFACT TYPES

ALBERT C. SPAULDING

WITHIN RECENT YEARS there appears to have been an increasing awareness on the part of archaeologists that certain statistical techniques offer economical methods of extracting information of cultural significance from archaeological data. The discussions of Kroeber (1940), Robinson (1951), and Brainerd (1951) have appeared in *American Antiquity*, and the last two even evoked a comment* (Lehmer, 1951). In addition to these papers, which are primarily devoted to exposition of method, a considerable number of special applications can be found in the literature. Archaeological research inevitably brings the researcher face to face with the problems of ordering and comparing quantities of data and of sampling error. There seems little doubt that the best approach to these problems involves a search of statistical literature for appropriate methods.

The discussion which follows is an attempt to apply certain statistical methods to the discovery and definition of artifact types and to suggest other applications to related problems. No effort has been made to explain such important statistical concepts as population, random sample, sampling error, and so on; these explanations are the proper function of textbooks, and any paraphrasing here would be presumptuous. I am indebted to Paul S. Dwyer, Consultant in the Statistical Research Laboratory of the University of Michigan, for reading and commenting on an earlier version of the manuscript.

The artifact type is here viewed as a group of artifacts exhibiting a consistent assemblage of attributes whose combined properties give a characteristic pattern. This implies that, even within a context of quite similar artifacts, classification into types is a process of discovery of combinations of attributes favored by the makers of the artifacts, not an arbitrary procedure of the classifier. Classification is further an operation which must be carried out exhaustively and independently for each cultural context if the most fruitful historical interpretations are to be made. It is the primary purpose of this paper to argue that with the aid of suitable statistical techniques the degree of consistency in attribute combinations can be discovered in any meaningful archaeological assemblage provided sufficient material is at hand, and hence that valid types can be set up on the basis of analysis of material from one component.

Wholesale acceptance of these views entails modification of a widely held concept of typology which has been clearly expressed by Krieger (Krieger, 1944; Newell and Krieger, 1949). Under this concept, the method employed to demonstrate the existence of a valid type is a site-to-site comparison to show consistency of the identifying pattern, range of variation, and historical relevance. In the absence of a method for investigation of consistency and range of variation within the site, this is indeed the only convincing technique available for validation of a proposed type. On the other hand, the presence of an adequate method for investigating consistency and range of variation within the site obviates a comparative study so far as the questions of the existence and definitive characteristics of a type are concerned. Historical relevance in this view is essentially derived from the typological analysis; a properly established type is the result of sound inferences concerning the customary behavior of the makers of the artifacts and cannot fail to have historical meaning. This is not meant to imply that corroborative evidence from the other sites would not be welcome in the case of a dubious type, i.e.,

* See pp. 341-53 in this issue for an application of the Robinson technique.

one which is on the borderline of probability owing to a deficient sample or lack of clear evidence of attribute clustering, nor is it meant to imply that the classifier is relieved of the responsibility of avoiding synonymy. Finally, it is not intended to assert that artifact types are the only useful units of attribute association for site-to-site comparison; numerous examples of good comparative work with body sherds are sufficient refutation of such an assertion, although the common practice of failing to distinguish between kinds of body sherds and types of vessels is a stumbling block in understanding the cultural meaning of a comparison. It should be pointed out that this discussion owes much to the expositions of Rouse (1939, especially pp. 9-23), Krieger (1944), Newell and Krieger (1949, pp. 71-74), and Taylor (1948, especially pp. 113-130).

The customary technique of classification consists of inspection and segregation of obtrusive combinations, or occasionally of attempting to describe all of the observed attribute combinations on an equal basis. Categories resulting from both of these methods are called "types," although they are not exactly comparable. Both methods fail to yield surely artifact types in the sense in which the term is used here. In the first case, segregation of obtrusive combinations, the cultural implications of the data are usually not exhausted, although under favorable circumstances all of the types may be discovered and described. In the second case, description of all combinations, the problem of typology is not faced at all; some of the "types" described will in all probability consist of combinations habitually avoided by the makers of the artifacts. Questions of typology arise, of course, only in a situation where a considerable variety exists within a group of generally similar artifacts — it is obvious that a stone projectile point and a pottery vessel belong in two separate artifact types. But within a group of similar artifacts the propriety of division into more than one type may be anything but obvious. It follows from the concept of the type adopted here that a pronounced association of two attributes is the minimum requirement for the demonstration of the existence of an artifact type, since two is the smallest number which can be considered an assemblage.

Application of this concept to concrete material can be illustrated by a few simple examples. Inspection of a collection of 100 vessels which represent all the pottery from a component results in the noting of the following attributes: smooth surface, cord wrapped paddle stamped surface, grit tempered paste, and shell tempered paste. The question to be answered is whether the vessels represent one or two pottery types with respect to these attributes. The data necessary to answer the question are the frequencies of vessels in each of the four possible categories into which two pairs of alternatives can be grouped, here smooth surface and grit temper, smooth surface and shell temper, cord wrapped paddle stamped surface and grit temper, and cord wrapped paddle stamped surface and shell temper. Table 7 presents these frequencies in 2 x 2 form under the assumption that the count in each category is 25 vessels.

TABLE 7. FOUR-CELL FREQUENCIES WITH NO ASSOCIATION OF ATTRIBUTES

	Grit Temper	*Shell Temper*	*Total*
Stamped Surface	25	25	50
Smooth Surface	25	25	50
Total	50	50	100

It is evident by inspection that the 100 vessels cannot be separated into two types under these circumstances. The cord wrapped paddle stamped vessels are equally divided with respect to grit temper and shell temper, the same is true of the smooth surfaced vessels, and conversely both the shell tempered and the grit tempered vessels are equally divided with respect to surface finish. A mathematical statement to the same effect can be obtained by applying the simple and useful four-cell coefficient of association described by Kroeber (1940). If the upper left cell is designated *a*, and the upper right cell *b*, the lower left cell *c*, and the lower right cell *d*, the coefficient of association for the attributes grit temper and cord wrapped paddle stamped surface would be computed as

$$\frac{(a+d)-(b+c)}{a+b+c+d} = \frac{(25+25)-(25+25)}{25+25+25+25} = \frac{0}{100} = 0.$$

The same result would follow for the other three pairs. The opposite situation would be that of Table 8. Here there are plainly two types with respect to the traits considered, a cord wrapped paddle stamped and grit tempered type and a smooth surfaced and shell

TABLE 8. FOUR-CELL FREQUENCIES WITH PERFECT ASSOCIATION OF ATTRIBUTES

	Grit Temper	*Shell Temper*	*Total*
Stamped Surface	50	0	50
Smooth Surface	0	50	50
Total	50	50	100

tempered type. The computed coefficient of association for the attributes grit temper and cord wrapped paddle stamped surface is

$$\frac{(50 + 50) - (0 + 0)}{50 + 0 + 50 + 0} = +1.0,$$

and the same coefficient would be obtained for the shell tempered, smooth surfaced category. On the other hand, the calculation for the smooth surfaced, grit tempered category shows

$$\frac{(0 + 0) - (50 + 50)}{0 + 50 + 0 + 50} = -1.0,$$

and this is also true of the cord wrapped paddle stamped, shell tempered category.

This discussion of four-cell coefficients has been introduced chiefly to illuminate the concept of the two attribute association as the minimum requirement for the establishment of an artifact type, although the simple four-cell coefficient of association and its more sophisticated relatives are by no means to be ignored as working methods under the proper conditions. One of the serious deficiencies of the four-cell coefficient is its failure to consider the vagaries of sampling, since a conservative interpretation of the material from any archaeological component requires that it be considered no more than a sample drawn from a universe of artifacts manufactured by a society over some vaguely defined period of time. Other precautions to observe when using four-cell coefficients are discussed by Kroeber (1940).

Methods do exist which give answers expressing the combined result of the error involved in sampling and the extent to which the observed data fit the expected with respect to a hypothesis. The remainder of this paper will be devoted to illustrating the application of these methods to typological problems and some other archaeological data. All of the techniques presented are drawn from the literature of biological statistics dealing with the analysis of binomial distributions, especially the discussions of Mather (1947, Chapter XI) and Snedecor (1946, Chapters 9 and 16), and the reader is referred to these sources for an adequate explanation of the underlying concepts. The most practical method of recording and subsequently extracting the variety and quantity of data needed for a thorough analysis of any sizable collection would appear to be one of the mechanically or electrically sorted punch card systems.

TABLE 9. FOUR-CELL FREQUENCIES WITH INDEPENDENCE OF ATTRIBUTES

	Grit Temper	*Shell Temper*	*Total*
Stamped Surface	53	64	117
Smooth Surface	32	43	75
Total	85	107	192

Using Table 9 as an example, an analysis which fulfills the stipulated conditions can be made by means of a formula for computing a statistical entity known as chi square. The formula which is most convenient for a 2 x 2 table is

$$\chi^2 = \frac{n(ad - bc)^2}{(a + b)\ (c + d)\ (a + c)\ (b + d)},$$

or verbally, the number of specimens multiplied by the squared difference of the product of the diagonals divided by the product of the marginal totals. Substituting the values of Table 9 gives

$$\chi^2 = \frac{192\ [\ (53 \times 43) - (35 \times 64)\ ]^2}{117 \times 75 \times 85 \times 107},$$

which reduces to

$$\frac{192 \times 231 \times 231}{117 \times 75 \times 85 \times 107} = 0.128.$$

With a χ^2 of 0.128 and one other argument, the number of degrees of freedom, it is possible to enter a table of χ^2 and read the probability of the occurrence of so large a χ^2 through the operation of sampling variation alone in a population having independent attributes in the ratios indicated by the marginal totals. The appropriate number of degrees of freedom is 1 because the computation imposes the restriction that the frequencies must add up to the marginal totals, so that as soon as a frequency is assigned to any cell those of the other three can be found by subtraction. The probability corresponding to a χ^2 of 0.128 with 1 degree of freedom is between .80 and .70, which means that a χ^2 this large would arise by chance alone between 70 and 80 times in 100 in a population having independent attributes.

It seems reasonable to accept the hypothesis of independence of attributes and conclude that the marginal totals present a fair picture of the potters' habits, there being very little evidence that the individual cell frequencies fall outside the range expected in a random drawing from a homogeneous population having the proportions of attributes indicated by the marginal totals. In other words, there is no discernible tendency for the attributes to cluster into types. Here, in contrast to the coefficients of association mentioned above, it has been possible to make a statement in terms of numerical probability and a definite hypothesis, which reduces the data to their most comprehensible form.

Chi square for Table 8 would be computed as

$$\frac{100\,[\,(50 \times 50) - (0 \times 0)\,]^2}{50 \times 50 \times 50 \times 50} = 100,$$

a value exceeding by a large amount the tabled value of 10.877 for a probability of .001 for 1 degree of freedom, and the probability that the marginal totals fairly represent the potters' habits is astronomically remote. The attributes are not independent; inspection of the table shows that the sample is derived from two populations, one characterized by grit tempering and a cord wrapped paddle stamped surface, the other by shell tempering and a smooth surface. This is the same conclusion as that based on the coefficient of association, but again a numerical expression of the odds against the occurrence of such a distribution in a random drawing from a population having an independent distribution of the four attributes has been provided.

It is important to note that the proportions used in testing attribute independence or lack of it were derived from the sample, and consequently the calculations have not tested the proposition that the observed proportions exactly represent those of the population from which the sample was obtained. What has been tested is the hypothesis that the two samples, those in the two rows or the two columns, were randomly drawn from a common binomial population. In the first instance (Table 9) the hypothesis was accepted, in the second (Table 8) it was rejected. Acceptance in the case of the data of Table 9 indicates that both cord wrapped paddle stamped and smooth surfaced vessels were randomly drawn from a population of vessels having grit temper and shell temper in a ratio estimated to be in the neighborhood of 85:107, or alternatively, both grit tempered and shell tempered vessels were randomly drawn from a population of vessels having cord wrapped paddle stamped and smooth surfaces in a ratio estimated to be 117:75. The estimated ratios are simply the marginal totals, and the inferences about the nature of the parent population can be completed by finding confidence limits for these estimates. This can be accomplished easily by means of a calculation or by reference to a table of confidence intervals such as that presented by Snedecor (1946, p. 4). Rejection of the hypothesis of independence in the case of Table 8 leads to the conclusion that cord wrapped paddle stamped vessels were drawn from a population of vessels estimated to be exclusively grit tempered, and smooth surfaced vessels were drawn from a probably exclusively shell tempered population. Again confidence intervals can be assigned to the estimates.

The next question to be investigated is that of a suitable technique for situations involving combinations of more than two pairs of attributes. The method to be employed is closely related to that just illustrated, but the resemblance is obscured by the streamlined computing routine used for the 2 x 2 table. There are two basic steps required: (1) calculation of an expected frequency for the combination, customarily under the hypothesis that the combination in question does not constitute a distinctive type, i.e., that the attributes making up the combination have independent distributions; and (2) comparison of the expected frequency with the observed frequency to determine whether or not the difference between the two can be reasonably attributed to sampling error. If the observed frequency exceeds the expected frequency by an amount too great to be considered the result of mere sampling error, it will be concluded that a genuine tendency for the makers of the artifacts to combine the attributes in question has been discovered — that the existence of a type has been demonstrated.

The following data will be used to explain the working method: in a collection of 297 pottery vessels, it is suspected that a combination of grit tempering, stamped surface, and a collared rim occurs often enough to provide sufficient grounds for the definition of a pottery type. A count made of the frequency of the triple combination gives 83 vessels; of the frequency of grit tempering alone, 117 vessels; of stamped surface alone, 91 vessels; and of collared rims alone, 136 vessels.

Under the hypothesis of independent distribution of attributes (no type), the frequency of the combination would be expected to be a simple function of the relative frequencies of the component attributes. Calculation of the expected number is a straightforward problem in compound probability, here

$$\frac{117}{297} \times \frac{91}{297} \times \frac{136}{297} \times \frac{297}{1} = 16.42 \text{ vessels.}$$

In practice it is necessary to compute the proportion (p) characteristic of the combination for reasons to be explained below. The computation of p here is

$$\frac{117}{297} \times \frac{91}{297} \times \frac{136}{297} = .0553.$$

The next step is to obtain the deviation (d) of the observation from the expectation by subtracting 16.42 from 83.00, which results in a deviation of 66.58.

It is necessary here to introduce some new symbols required for the final comparison of the expected frequency (E) and the observed frequency (O). The proportion of vessels not expected to exhibit the combination will be designated q, which is simply $1-p$ or $1.000-.0553=.9447$ in the example. The expectation for the various possible frequencies of two alternative types (in this example grit tempered, cord wrapped paddle stamped, collared rim vessels and vessels not having this combination) can be found by expanding the binomial $(p+q)^k$, where k is the symbol for the number of individuals in the group (297 vessels); in addition, and of immediate importance in the solution of the problem, is the fact that the variance of the expanded binomial distribution is pqk ($.0553 \times .9447 \times 297 = 15.52$). The standard deviation (σ) is $\sqrt{pqk}$, which makes it possible to compute easily either the deviate in units of standard deviation as $\frac{d}{\sigma}$, or $\left[\frac{d}{\sigma}\right]^2$ as $\frac{d^2}{pqk}$. Both $\frac{d}{\sigma}$ and $\left[\frac{d}{\sigma}\right]^2$ can be converted into statements of probability by means of widely available tables. In the case of d/σ, tables of areas of the normal curve or tables for t for infinite degrees of freedom may be used; $(d/\sigma)^2$ is the familiar χ^2 for 1 degree of freedom. Choice of formula is a matter of individual preference since the answers obtained are identical; tables for χ^2 are less closely computed than those for d/σ owing to their two dimensional character, but the precision of the latter does not appear to have any advantage for archaeological purposes. In both cases the tables were computed on the basis of a continuous curve rather than the binomial curve with discrete steps used here, and consequently they are not exactly applicable. A widely recommended procedure for avoiding excessive distortion is to group categories so that the expected numbers are not too small, say 5 or less. A partial correction (the Yates correction) can be made by adjusting d, and precise methods of adjustment for small numbers can be found in statistical literature. The simple adjustments do not seem to change the results markedly, but anyone planning to use these techniques should be familiar with informed discussions of the subject.

Calculations for $(d/\sigma)^2$ for the example are

$$\frac{(66.58)^2}{15.52} = \frac{4432.90}{15.52} = 285.62.$$

Entering a table of χ^2 with this figure and 1 degree of freedom, a probability of finding a fit with hypothesis through chance at least as bad of very much less than .001 is noted. A similar calculation for d/σ indicates that the odds are actually less than 1 in 400,000,000,000 that so large a difference between observed and expected frequencies would arise through random sampling in the expanded binomial. It can be concluded that the chance of a sampling vagary as the explanation is exceedingly remote, and the large number of vessels exhibiting the combination must be attributed to the habits of the potters. The calculation does show that a pottery type exists. Further research would be necessary to investigate whether (1) on the basis of other attributes it might not be possible to identify a group of pottery types sharing the specified combination, or (2) whether there are other combinations differing by only one attribute which should be included in the type description as variants. The original conclusion — that the existence of a pottery type was demonstrated — is not modified by either case.

The evaluation of probability can perhaps be clarified by two other examples. Had the observed frequency been 24 vessels, χ^2 would have been computed as

$$\frac{(7.58)^2}{15.52} = \frac{57.46}{15.52} = 3.70,$$

which for 1 degree of freedom represents a probability of between .10 and .05, but much closer to .05. The conclusion is not at all clear. There is an appreciable chance that no real preference for the combination was exhibited by the potters, and the evaluation must be made with the aid of all the experience which the archaeologist can muster. If related sites plainly show that the combination is elsewhere a valid type, the interpretation would probably be that in this case the type was just appearing or disappearing. In the absence of other data, one could say only that there is a very good possibility that a type has been discovered. In certain types of statistical investigation a χ^2 of more than 3.841 (the .05 level of probability for 1 degree of freedom) is considered significant, or in our terms the hypothesis of independence would be rejected. It would appear unwise to carry over blindly such concepts into archaeology. Had the observed frequency of the combination been 8 vessels, d would have been 8.42 and $\chi^2 = 4.57$ with a probability between .05 and .02. The same general reasoning applies again, but here the situation is reversed because the expected frequency exceeds the observed frequency; there is a strong probability that the potters tended to avoid the combination, and the examples observed might best be considered the work of unorthodox potters.

A thorough investigation of a collection requires the calculation of d/σ or χ^2 for every possible combination of presumably important attributes. The number of combinations possible can be found by grouping the mutually exclusive attributes and multiplying together the number of attributes in each of the groups. If the groups of attributes consist of (1) smooth surface, stamped surface; (2) incised rim, plain rim; (3) incised lip, plain lip; and (4) bowl shape, jar shape; the computation is 2 x 2x 2 x 2 = 16 possible combinations. If the groups are (1) smooth surface, stamped surface; (2) rectilinear incising on shoulder, curvilinear incising on shoulder, plain shoulder; and (3) grit tempered, shell tempered; there are 2 x 3 x 2 = 12 possible combinations. These 12 combinations will be used in an example with the following data given: total number of vessels (k), 186; frequency of smooth surface, 121 vessels; of stamped surface, 65 vessels; of rectilinear pattern incised on shoulder, 47 vessels; of curvilinear pattern incised on shoulder, 28 vessels; of plain shoulder, 111 vessels; of grit tempering, 70 vessels; and of shell tempering, 116 vessels. Combination counts and computations are shown in Table 10. The computations are exactly like those described above. For example, p in the first combination is

$$\frac{121}{186} \times \frac{47}{186} \times \frac{70}{186} = .0619.$$

The expected number (E) is 186 x .0619 = 11.51, and so on.

TABLE 10. COMPUTATION OF $\frac{d^2}{pqk}$ FOR TWELVE COMBINATIONS OF ATTRIBUTES

Attribute Combination	O	E	d	d^2	pqk	$\frac{d^2}{pqk}$
Sm. surf., rect. sh., grit t.	0	11.51	−11.51	132.02	10.78	12.25
Sm. surf., curv. sh., grit t.	2	6.84	−4.85	23.52	6.59	3.56
Sm. surf., plain sh., grit t.	14	27.17	−13.17	173.45	23.21	7.47
Sm. surf., rect. sh., shell t.	38	19.07	+18.93	358.35	17.11	20.94
Sm. surf., curv. sh., shell t.	26	11.36	+14.64	214.33	10.66	20.11
Sm. surf., plain sh., shell t.	41	45.04	−4.04	16.24	34.13	0.48
St. surf., rect. sh., grit t.	3	6.18	−3.18	10.11	5.97	1.69
St. surf., curv. sh., grit t.	0	3.68	−3.68	13.54	3.61	3.75
St. surf., plain sh., grit t.	51	14.60	+36.40	1324.96	13.45	98.51
St. surf., rect. sh., shell t.	6	10.25	−4.25	18.06	9.69	1.86
St. surf., curv. sh., shell t.	0	6.10	−6.10	37.21	5.90	6.31
St. surf., plain sh., shell t.	5	24.20	−19.20	368.26	21.04	17.50
Total	186	186.00	0.00			

Table 10 is to be interpreted simply as a list of χ^2 values, each of which has its corresponding probability for 1 degree of freedom. The individual χ^2 values, computed as d^2/pqk, do not have additive properties in contrast to the contingency table discussed below. Interpretation in terms of pottery types follows the principles already discussed. Three combinations have large positive deviations and large χ^2 values with probabilities well beyond the .001 level. These are stamped surface, plain shoulder, grit temper; smooth surface, curvilinear incised shoulder, shell temper; and smooth surface, rectilinear incised shoulder, shell temper. The last two combinations differ by only one attribute, and hence are to be lumped in one type. The same is true of the smooth surfaced, plain shouldered, shell tempered combination, which is important numerically but has a very small χ^2 value. Accordingly, there is definitely a smooth surfaced, shell tempered type having three kinds of shoulder treatment in a ratio estimated to be about 26:38:41. This can be confirmed by calculating a χ^2 for a 2 x 2 table testing the degree of association of smooth surface and shell temper. It will be found that they are very strongly associated, as are grit temper and a stamped surface. It can be inferred that the indifferent χ^2 value (0.48) of the shell tempered, plain shouldered, smoothed surface combination is the result of the fact that plain shoulders are shared with and are rather more characteristic of the stamped surfaced, grit tempered combination. This conclusion is at sharp variance with conventional type analysis, where the shell tempered, plain shouldered, smooth surfaced combination would almost surely be distinguished as a separate type, as would the other two smooth surfaced, shell tempered combinations. The calculations above are intended to be an objective demonstration that the fundamental pattern of the type is the smooth surfaced, shell tempered vessel. Shoulder treatment can be described only in terms of estimated ratios of a group of mutually exclusive attributes.

The stamped surfaced, plain shouldered, grit tempered vessels constitute a second definite type; χ^2 for the combination is very high (98.51) and it can be shown that stamped surface and grit temper are strongly associated. The 14 vessels having smooth surfaces, plain shoulders, and grit temper would not be assigned to either type; they are genuinely intermediate and would be so described. The same reasoning applies to the 5 vessels having stamped surfaces, plain shoulders, and shell tempering. The remaining few vessels share two attributes with one or the other of the types and would be assigned accordingly as somewhat aberrant examples. Combinations of this sort, characterized by negative deviations and crossing over of attributes from two types, offer interesting evidence on the degree of conventionality of the potters. In this connection the combinations with a frequency of 0 are highly informative.

A second sort of table can be computed which offers summary evidence on the total pottery making habits of the group. For this table, the individual contribution of each combination would be computed as d^2/E, which for the first combination of Table 10 is 132.02/11.51. The total of these contributions is a χ^2 value for the 12 combinations taken together, for which a probability can be found in the χ^2 table using 7 degrees of freedom. A verbal explanation of the appropriateness of 7 degrees of freedom is too cumbersome for inclusion here, and a clear graphic presentation of a 2 x 3 x 2 table is also difficult, but it can be stated that the particular restrictions imposed by the attribute totals used as basic data allow 7 of the 12 cells of the table to be filled in freely within the general limitations of the attribute totals. The remaining five can be determined by subtraction and hence do not contribute to the degrees of freedom. A χ^2 computed in this manner gives an over-all measure of the tendency of the potters to group attributes and offers cogent material for comparison with other sites having the same categories. Other sorts of comparisons between sites can be made by using the observed number for each combination from one site as the expected number for the other and calculating the resulting χ^2 or by calculating a χ^2 testing the proposition that both sets of observed values could reasonably be considered random samples from a common population. The latter process is illustrated below in the example dealing with the problem of site homogeneity (Table 12).

All of the examples have been concerned exclusively with attributes which are physical properties of the artifacts. It is well known, however, that artifacts have other kinds of attributes, notably provenience, which can be pertinent evidence for the existence of a type. Thus a site might yield two kinds of vessels which differed only in the presence or absence of a single physical attribute, say a lip flange on one. If nothing but physical properties were considered, both kinds would be included in one pottery type because a difference of one attribute is not sufficient evidence for separation. But if the flanged lip appeared only on vessels found in graves and the plain lip was confined to village debris, it would be obvious that the potters had in mind two types with different functional connotations. Provenience furnishes the second attribute required to differentiate two types. The attributes "found in graves" and "found in village refuse" can be included in a probability calculation in exactly the same way as can any physical property of an artifact.

An example, this time not fictitious, of the application of this technique to a non-typological problem will be presented. The data of Table 11 are from the Columbia University excavations at the Arzberger Site, Hughes County, South Dakota, and summarize provenience data of grooved paddle stamped body sherds and other types of surface finish. The

TABLE 11. SURFACE FINISH OF BODY SHERDS BY PROVENIENCE, ARZBERGER SITE, SOUTH DAKOTA

Excavation Unit	*Surface Finish*		
	Grooved Paddle Stamped	*Other*	*Total*
House I	396	1,279	1,675
House II	135	546	681
House III	172	532	704
House IV	178	657	835
Ditch	0	4	4
Unknown	22	79	101
Total	903	3,097	4,000

problem to be investigated is one of site homogeneity. If the site is homogeneous, one excavation unit should be much like another within the limits of sampling error. With respect to the data given on surface finish of body sherds, a hypothesis of independence can be set up: the proportion of grooved paddle stamped sherds will be a function of the frequency of the totals and will be independent of the locus from which the sample is drawn if the site is truly homegeneous. Chi square is computed by the d^2/pqk method used above, although this is not the most common technique for a $2 \times n$ contingency table such as is given. The value of p is $903/4{,}000 = .2258$, $q = .7742$, and k is successively the total number of sherds for each sample. The values are shown in Table 12 (a few rounding errors have not been adjusted). The result is good evidence that the

TABLE 12. TEST OF HOMOGENEITY OF EXCAVATION UNITS, ARZBERGER SITE, SOUTH DAKOTA

	O	E	d	d^2	pqk	$\frac{d^2}{pqk}$
House I	396	378.21	17.79	316.48	292.81	1.08
House II	135	153.77	18.77	352.31	119.05	2.96
House III	172	158.96	13.04	170.04	123.07	1.38
House IV	178	188.54	10.54	111.09	145.97	0.76
Other[1]	22	23.70	1.70	2.89	18.35	0.16
Total	903	903.00				$\chi^2 = 6.34$

[1] The expected frequency for "Ditch" is less than 6, and accordingly it is incorporated in a new category by adding its value to "Unknown."

hypothesis of independence is correct. Individual values are small, and the total for 4 degrees of freedom (this is a 2 x 5 contingency table) corresponds to a probability of between .20 and .10, which does not give any very convincing reason to suspect significant differences in the various excavation units. It can be concluded that so far as the evidence at hand is concerned, the site may reasonably be considered the product of a single occupation over a restricted period of time.

An attempt to appraise the usefulness of this approach to typological and related problems should consider the amount of labor necessary in making the computations. In view of the general availability of computing machines, this seems trivial. The writing of the exposition was far more tedious than the computing of the examples. There is a great deal of work required in making, recording, and assembling the observations needed for a thorough study, but this is not the fault of the statistical methods. It is rather an inevitable part of any detailed study. The methods of calculation used here were selected on a basis of clarity of exposition, not economy of labor; those interested in computing routine are referred to the statistical textbooks cited.

With regard to the more serious question of general usefulness, these are the methods generally recommended for handling data of this sort, although no claim is made that the particular procedures illustrated here completely exhaust the resources of statistics. The information derived from them is important in an earnest attempt to discover the cultural significance inherent in archaeological remains, and there is no other way in which such information can be obtained. There is no magic involved, however; the usefulness of the result is entirely dependent upon the wisdom with which attributes are observed and investigated and on the relevance of the context to meaningful archaeological problems. Moreover, the inference to be drawn from a statement of probability is sometimes not altogether clear, but at least the degree of uncertainty is put into objective form.

A source of uncertainty which has been mentioned is the fact that the proportions on which the hypothesis of independence is evaluated are derived from the sample and hence are themselves subject to sampling error. This difficulty is inescapable; we can work only with the samples we have, and the observed proportions are surely the best estimate of the proportions of the population, the properties of which must be inferred from the sample. Nevertheless, the cautious student will interpret his results with one eye on a table of confidence limits. To add to this uncertainty, the dimensions of which can at least be estimated on the basis of statistical theory, there is the purely archaeological problem of the nature of the relationship of the sample to the living culture which produced the artifacts. The whole problem is summarized by the often repeated warning that statistics are never a substitute for thinking. But statistical analysis does present data which are well worth thinking about.

BIBLIOGRAPHY

BRAINERD, GEORGE W.
1951. The Place of Chronological Ordering in Archaeological Anaylsis. *American Antiquity,* Vol. 16, No. 4, pp. 301-13. Salt Lake City.

KRIEGER, ALEX D.
1944. The Typological Concept. *American Antiquity,* Vol. 9, No. 4, pp. 271-88. Menasha.

KROEBER, A. L.
1940. Statistical Classification. *American Antiquity,* Vol. 6, No. 1, pp. 29-44. Menasha.

LEHMER, DONALD J.
1951. Robinson's Coefficient of Agreement — A Critique. *American Antiquity,* Vol. 17, No. 2, p. 151. Salt Lake City.

MATHER, K.
1947. *Statistical Analysis in Biology.* New York: Interscience Publishers.

NEWELL, H. PERRY, AND ALEX D. KRIEGER
1949. The George C. Davis Site, Cherokee County, Texas. *Memoirs of the Society for American Archaeology,* No. 5. Menasha.

ROBINSON, W. S.
1951. A Method for Chronologically Ordering Archaeological Deposits. *American Antiquity,* Vol. 16, No. 4, pp. 293-301. Salt Lake City.

ROUSE, IRVING
1939. Prehistory in Haiti, A Study in Method. *Yale University Publications in Anthropology,* No. 21. New Haven.

SNEDECOR, GEORGE W.
1946. *Statistical Methods Applied to Experiments in Agriculture and Biology.* Ames: Iowa State College Press.

TAYLOR, WALTER W.
1948. A Study of Archaeology. *Memoirs, American Anthropological Association,* No. 69. Menasha.

Museum of Anthropology
University of Michigan
Ann Arbor, Michigan
August, 1952