Statistical Analysis in Archaeology

Edward V. McMichael, Indiana University

The purpose of this paper is to determine the extent of use of statistical methods in archaeological analysis. In the broad sense of the word, statistics can be used almost synonymously with quantification; such usage in archaeological analysis is ubiquitous. Here the usage will be restricted to "sophisticated" statistics, which in practice will amount to the use of correlation coeficients or the use of reliability estimates.

Certain exclusions and restrictions will be observed *a priori*. In way of applications, carbon-14 standard deviations will not be considered since this usage of the statistical method stems from another discipline entirely and is not representative of archeological interests in statistics. Also any use of standard deviation or other statistical devises in connection with bio-anthropological data authored by physical anthropologists is to be excluded for the same reason. The area of interest will be restricted to the Americanist field simply as a practical consideration.

The "statistics" this paper will deal with fall into two categories, correlations and reliability measures. The correlations encountered were all of the attribute type, that is, two by two classifications with the correlation figure indicating the extent of co-occurrence. These correlations may be of several forms, e.g., Phi, Q, Z. Each of these has its advantages and disadvantages, and it is well to investigate the characteristics of each before using any of these correlations. A reliability measure on the other hand is concerned with the probability that a significant difference exists between two or more samples. Reliability measures encountered are: Chi Square (X^2), t, and the Analysis of Variance (F ratio).

To achieve a representative sample all of the articles in *American Antiquity* were examined. The sum total of these articles represents the trends in Americanist archaeology and among American archaeologists for the present purposes, since it is the only professional journal in the United States devoted to American archaeology. Also the period of the journal's existence approximates the mature period of American archaeology (1935 to date). Thus, Volumes I through XXIII were examined (1935-1936 to 1957-1958).

The precise method used in this analysis was to leaf through each volume of *American Antiquity* seeking any "sophisticated" statistical applications. The results of this procedure will be presented, in chronological order, with a brief résumé of the methods used by each author.

In Volumes I through V (1935-1940) no articles containing statistical applications of any kind appear. The first such article is in 1940 by Kroeber in which he lucidly discusses the merits and demerits of the use of statistics in archaeological classification, and applies several different

correlation coefficients to previously presented archaeological data, in two by two box form (6). The following coefficients are used:

$$\text{Phi} : \frac{ad-bc}{\sqrt{(a+b)\ (a+c)\ (b+d)\ (c+d)}} \quad \text{or} \quad \sqrt{\frac{X^2}{N}}$$

$$Q : \frac{ad-bc}{ad+bc}$$

$$Z : \frac{(a+d) - (b+c)}{N}$$

$$W : \frac{a+d}{N}$$

Unfortunately, Kroeber becomes enamoured of Z and W whose possibilities and limitations have not been fully explored. In general, Kroeber would make more application of statistics in archaeology, but he has reservations, and points out a number of pitfalls such as the "personal equation" in description and typology. His statistical work in this paper confirms the previous investigators' non-statistical conclusions, though in some cases minor divergences emerge.

Fairbanks in Volume VII, uses Kroeber's Z to compare the traits of several southeastern shell mound cultures (4). This analysis is directly inspired by Kroeber's article of the previous year, and again subjective impressions are largely borne out by the statistical application. Kroeber in the same Volume returns to statistics to further analyze the data of another author (7). Again he uses his Z coefficient and to good effect. These earliest uses of statistics in archaeology can all be laid at the feet of Kroeber and it is understandable in the context of the University of California's "statistical phase" of the 1930's and early 1940's when Driver, Klimek and Kroeber were all applying statistics prodigiously in ethnology. The spurt in the application of statistics to archaeology is obviously inspired from this source.

Unfortunately, these few early attempts largely went unnoticed or were pointedly ignored. It is not until 1947 (Volume XIII) that statistical methods again appear in *American Antiquity* in an excellent article by Cook and Treganza in which they chemically analyze the constituents of two shell mounds (3). They wisely apply some reliability checks to their work. This is principally what they call a "critical ratio," which is an equivalent to t, the formula being:

$$\text{C. R.} : \frac{\text{difference between means}}{\sqrt{\text{sigma}_1\ \text{sigma}_2}} \quad \text{or} \quad \frac{X_1-X_2}{\sqrt{\sigma_1\ \sigma_2}}$$

Also, Chi Square (X^2) and standard deviations are introduced in the process of macro- and micro-analysis of the mounds.

Again it is several years before any renewed interest is manifested in statistics. Then in 1951, Robinson and Brainerd introduce their method for the chronological ordering of archaeological deposits and the Coefficient of Agreement (2, 10). The Coefficient in formula is:

$200—(\Sigma p_1 - p_2)$ where p_1 is the percentage expression of a type within one unit

p_2 is the percent of the same type in the unit to be compared.

The resultant coefficients are tabled and clustered with the highest coefficients closest to the diagonal. This provides the best chronological ordering of the data, barring unusual circumstances. This technique inspired Lehmer (9) to suggest the use of a mean standard error in the coefficient calculation to minimize differences in sample size, but Robinson and Brainerd adequately expose the fallaciousness of such a procedure (11).

Spaulding makes a significant advance in his "Statistical Techniques for the Discovery of Artifact Types" (12). He advocates the use of correlation, citing Kroeber's Z, to discover artifact types. Spaulding then discusses problems of sampling and introduces several varieties of X^2 as a measure of reliability. This article in general is quite excellent; one only wishes for more development of his methods with extended practical applications. Spaulding's paper in turn stimulated Ford to write a brief comment to which Spaulding replied, but this exchange was more concerned with culture theory than statistics (5, 13).

In the same year (1953), Belous reevaluates the Central California culture sequence which Heizer had established by impressionistic methods. Belous employs the Robinson-Brainerd method but uses non-ceramic traits whereas Robinson originally worked with pottery. Again the statistical analysis largely confirms the original non-statistical sequence.

In 1954, Laughlin and Marsh apply the analysis of variance, a reliability measure, to archaeological data (8). This measure attempts to state whether there is a significant difference between the means of samples, in this case the thickness of Lamellar flakes from the Aleutians. The resultant F ratio is significant for the whole series, but adjacent means are non-significant.

The last article included in the designated time span of *American Antiquity* is again by Spaulding in which he discusses several methods of testing the significance of difference for carbon-14 dates (14). With only two dates t is used, and for the example cited, a significant difference is noted. For more than two dates from a supposedly common source, analysis of variance is employed, and in the instance cited, no significant difference is found. Spaulding provides an eminently useful method applicable to carbon-14 dating.

The above articles are summarized in the following table. The volume and year of *American Antiquity* and the author of any statistical article is given; the statistical devices used are indicated (C in parentheses denotes a correlation coefficient, R indicates a reliability measure); and finally "Frequency" gives the individual number of correlations or reliability checks worked by each author to provide a further index to statistical activity in archaeology.

In summary then, the use of statistical methods in archaeological analysis can largely be reduced to a brief flurry caused by Kroeber in the early 1940's, followed in the early 1950's by the Robinson-Brainerd method of chronological ordering, and then by Spaulding's statistical contributions. Generally, statistical applications as seen in *American Antiquity* have been sporadic, but with some increase in frequency of articles concerned with the subject in later volumes.

TABLE 1

Articles in *American Antiquity*, vols. I-XXIII, containing statistical applications.

Vol. and Year	*Articles*	*Statistic*	*Freq.*
I-V, 35-40	none		
VI, 40-41	Kroeber	Q (C)	55
		Z (C)	61
		W (C)	10
VII, 41-42	Fairbanks	Z (C)	6
	Kroeber	Z (C)	21
VIII-XII, 42-47	none		
XIII, 47-48	Cook and		
	Treganza	t (R)	9
		X^2 (R)	1
XIV, XV, 48-50	none		
XVI, 50-51	Robinson-		
	Brainerd	Coef. of (C) Agreement	185
XVII, 51-52	Lehmer	none	
XVIII, 52-53	Spaulding	Z (C)	3
		X^2 (R)	17
	Belous	Coef. of Agreement (C)	120
	Brainerd-		
	Robinson	none	
XIX, 53-54	none		
XX, 54-55	Laughlin and		
	Marsh	F (R)	1
XXI, XXII, 55-57	none		
XXIII, 57-58	Spaulding	t (R)	1
		F (R)	1

The amount of statistical application in archaeology is pathetically small considering the tremendous possibilities in a field requiring so much quantification. Furthermore, all the articles described, except the one by Laughlin and Marsh (8) are just faltering initial steps, more concerned with giving examples than actually applying statistics to archaeological data as a matter of course.

Archaelogists are prone to make much ado about the scientific nature of their field. They proceed with utmost caution and care in their field excavations, making a virtual fetish of "Scientific methods." And then what? They proceed to analyze the data thus recovered in antiquated and subjective fashion, ignoring the possibilities of up-to-date quantitative analysis, namely statistics and allied methods. And just why is this? For the most part it seems to reside in the archaeologist's lack of knowledge and understanding of statistics. This lack is usually covered up by defensive arguments against the use of statistics, such as "statistics dehumanizes archaeology and reduces everything to numbers" a most absurd argument in view of the already dehumanized nature of a good portion of archaeology. Statistics is a means not an end; the results must be interpreted just as in any other type of analysis. In short, statistical analysis need be no more dehumanized

than any other type of analysis. Another argument may be called the "much ado about nothing" position; that after all is said and done in statistical analysis, one is no more sure of his results than before. But proponents of this line of argument merely demonstrate a lack of understanding of scientific method; science proceeds by successive approximations of the truth and at least statistical methods provide a more objective method of approaching an answer than the more commonly employed subjective approach. Even though the statistical analysis reported in these several articles in *American Antiquity* merely substantiate subjective conclusions, it should be emphasized that minor differences were frequently noted, and it is here, in uncovering minor nuances, that one value of statistical analysis lies.

Enlightened application of statistics to problems in archaeology has much to offer the field; indeed the accumulation of more and more archaeological data will make statistical analysis more and more a necessity, not just a convenience. This is not to say that everything must be analyzed statistically; this would probably be as bad or worse than no statistics. But incorporation of statistical methods into archaeological analysis should be regarded as another tool to be used on appropriate occasions. Such use would enhance the value of archaeological analysis considerably and would greatly strengthen the conclusions of the field.

Literature Cited

1. BELOUS, RUSSELL E. 1953. The central California chronological sequence re-examined. American Antiquity **XVIII**:341-353.
2. BRAINERD, GEORGE W. 1951. The place of chronological ordering in archaeological analysis. American Antiquity **XVI**:301-313.
3. COOK, S. F. and A. E. TREGANZA. 1947. The quantitative investigation of aboriginal sites: comparative physical and chemical analysis of two California Indian mounds. American Antiquity **XIII**:135-141.
4. FAIRBANKS, CHARLES H. 1942. The taxonomic position of Stalling's Island, Georgia. American Antiquity **VII**:223-231.
5. FORD, JAMES A. 1954. Comment on A. C. Spaulding's 'Statistical Techniques for the Discovery of Artifact Types." American Antiquity **XIX**:390-391.
6. KROEBER, ALFRED L. 1940. Statistical classification. American Antiquity **VI**:29-44.
7. KROEBER, ALFRED L. 1942. Tapajo Pottery. American Antiquity **VII**:403-405.
8. LAUGHLIN, WILLIAM S. and GORDON H. MARSH. 1954. The lamellar flake manufacturing site on Anangula Island in the Aleutians. American Antiquity **XX**:27-39.
9. LEHMER, DONALD J. 1951. Robinson's coefficient of agreement—a critique. American Antiquity **XVII**:151.
10. ROBINSON, W. S. 1951. A method for chronological ordering archaeological deposits. American Antiquity **XVI**:293-301.
11. ROBINSON, W. S. and GEORGE W. BRAINERD. 1952. Robinson's coefficient of agreement—a rejoinder. American Antiquity **XVIII**:60-61.
12. SPAULDING, ALBERT C. 1953. Statistical techniques for the discovery of artifact types. American Antiquity **XVIII**:305-313.
13. SPAULDING, ALBERT C. 1954. Reply to Ford. American Antiquity **XIX**:391-393.
14. SPAULDING, ALBERT C. 1958. The significance of difference between radiocarbon dates. American Antiquity **XXIII**:309-311.