Statistics in Archaeology and Its Application to Ancient Near Eastern Data

Marilyn KELLY-BUCCELLATI and Ernestine S. ELSTER – Los Angeles

The increasingly widespread use of statistics as a tool in archaeological research is the motivation for this study. However all statistical models and their possible uses are not covered in this short article. Seriation techniques for purposes of relative dating are discussed and the application of statistical description and analysis to archaeological problems is outlined. Recent literature concerned with a wide variety of statistical methods including the relevance of computer technology can be found in Whallon (1972), Hodson, Kendall and Tãutu (1971), and Gardin (1970).

Vertical Patterning: Seriation

Definition and historical background

A fundamental problem frequently encountered by archaeologists is the establishment of a relative chronology for sites or groups of artifacts with no continuous stratigraphic relationship. One solution of chronological ambiguity is the isolation of distinct, important artifacts, types, or classes and the comparison of levels in which they are found in several sites or over entire regions. This results in an interconnected chronological framework of relationships (e. g. Kantor 1965; Dyson 1968; Gimbutas 1970). Among the sites some chronological evidence is given by stratigraphic position; other temporal indications come from outside sources such as texts and radiocarbon dating.

Another approach to this problem is the use of seriation, the ordering of artifacts in their presumed chronological sequence through the observance of their relative frequencies. Seriation is one of the most useful of the statistical tools dealing with chronological ambiguity. For example, to seriate sites in one region, specific categories of artifacts are chosen and their relative frequencies among the sites are computed. The assumption is that within a single geographical area sites with similar frequency distributions are contemporary (Deetz 1967: 26-30). On this basis the sites are then seriated. This technique may be used when the chronological associations are unknown and thus is applicable, for example, to material from a cemetery where the stratigraphy does not indicate the relative chronology of the graves or to sites within one region if the material is sufficiently homogeneous. The validity of seriation is even greater, obviously, if combined with stratigraphic information from at least some of the sites.

The use of seriation is not new, and in fact was first used on archaeological material from Egypt at the end of the 19th century. At that time Sir Flinders Petrie attacked the problem of chronologically ordering over 4000 graves from the cemeteries of Abadiyeh and Hu near Dendereh (Petrie 1899, 1901). Since the prehistoric sequence in Egypt was then unknown, Petrie established a chronological framework by using his own material exclusively. On the evidence of stylistic and technological differences, nine types of pottery from 900 tombs were analyzed. The results showed that certain variables interacted in a consistent fashion and from these his chronological "Sequence Dating" (S.D.) was developed. He divided the graves into fifty groups of eighteen tombs each, with each unit approximately succeeding the other in time. One sequence date, between S.D. 31-80, was assigned to each of these groups, thereby linking his artificial time to a projected population size. As in any seriation, Petrie had no way of determining from his ceramic evidence the chronological direction of the evolution of his sequence and could only confirm his assumptions through ties with Proto-Dynastic material.

Petrie's study did not take into account the problem of spacial variation among sites and regions. Thus, contrary to his expectations, its application to prehistoric sites in Lower Egypt and Nubia is unsuccessful (Massoulard 1949: 61-9). The necessity to evaluate this dimension was appreciated by Kroeber and Spier in their work in the American Southwest (Kroeber 1916; Spier 1917).

In an important study, David G. Kendall attempted to formulate the statistical problems involved in Petrie's approach. He also pointed out the contacts Petrie had with his colleague, Karl Pearson at University College, London (Kendall 1963). Pearson was interested in applying mathematical and statistical models to other disciplines. This was the first in what is by now an extensive literature by statisticians and mathematicians on the application of and theory behind seriation (Whallon 1972: 42-45). The techniques continued to be refined and tested in the first half of this century (Sterud 1967).

Techniques of seriation

A suitable statistical procedure to determine the relative correlation of types among various sites was designed by Robinson with the help of the archaeologist Brainerd (Brainerd 1951; Robinson 1951). Their approach places data, for which the spacial distribution is limited to a particular site (i.e. burial finds), or area (i.e. survey or excavated material), along a continuum of similarity. The aim, as with all seriation, is to establish relative temporal relationships.

In this method a coefficient of agreement is calculated between each pair of sites, thus giving numerical expression to the measure of difference between two sites. These coefficients are then ordered in a symmetrical matrix bringing the coefficient with the highest magnitudes along the diagonal of self agreement. To illustrate this method, we compare cylinder seals from Tchoga Zanbil found in Chapels III and IV as well as those from the Palace-hypogeum (Porada 1970). These three deposits cannot be dated from the stratigraphy so their chronological position is established by Porada on stylistic grounds (Porada 1970: 127-131). We test here her groups I, II, III, VII, XI and XIII. The total sample comprises 92 seals: 29 in Chapel III, 55 in Chapel IV and 8 in the

Palace. In the construction of the matrix, the initial step is to calculate the percentages of each group in each location.

	CH III	Ch IV	Pal
Gp I	10	20	0
Gp II	7	9	0
Gp III	21	5	0
Gp VII	31	38	12.5
Gp XI	3	13	12.5
Gp XIII	28	15	75
	100%	100%	100%

These percentages are then compared, two sites at a time, by computing the index of disagreement (ID) between them.

	Ch III	Ch IV	ID
Gp I	10	20	10
Gp II	7	9	2
Gp III	21	5	16
Gp VII	31	38	7
Gp XI	3	13	10
Gp XIII	28	15	13
			58 TOTAL ID

	Ch III	Pal	ID
Gp I	10	0	10
Gp II	7	0	7
Gp III	21	0	21
Gp VII	50	12.5	37.5
Gp XI	6	12.5	6.5
Gp XIII	44	75	31
			113 TOTAL ID

	Ch IV	Pal	ID
Gp I	20	0	20
Gp II	9	0	9
Gp III	5	0	5
Gp VII	38	12.5	25.5
Gp XI	13	12.5	.5
Gp XIII	15	75	60
			120 TOTAL ID

The coefficient of agreement is obtained by subtracting the ID from 200, which is the maximum amount of agreement to be seen between two sites. A 200 figure indicates that the sites are identical and zero would be the maximum disagreement score. In each of our examples the coefficient of agreement is: Ch III-Ch IV:142; Ch III-Pal:87; Ch IV-Pal:80. Next, these coefficients of agreement are placed unordered in a matrix.

	Ch III	Ch IV	Pal	
Ch III	200	142	87	
Ch IV	142	200	80	
Pal	87	80	200	Diagonal of Agreement

From this simple matrix, the other possible variations can be seen quickly: Ch III-Pal-Ch IV; Pal-Ch III-Ch IV. The best matrix, shown below, is the last:

	Pal	Ch III	Ch IV
Pal	200	87	80
Ch III	87	200	142
Ch IV	80	142	200

The chronological order indicated in this matrix is either Palace-Chapel III-Chapel IV or Chapel IV-Chapel III-Palace since the chronological direction is not given by the seriation.

The disadvantages of this sample are: (1) the typology was not designed for this type of study; (2) the Palace-hypogeum had few seals. Nevertheless, the seriation confirms the opinions of both Porada and Ghirshman as to the chronological sequence: Chapel IV, Chapel III, Palace (Porada 1970: 129).

In practice, the ordering of any matrix is a long, tedious task, and computer programs have been developed which reduce the required time (Ascher and Ascher 1963; Kuzara, Mead, Dixon 1966; Hole and Shaw 1967; and current work discussed by Whallon 1972).

Another technique also aimed at finding the best seriated order using a correlation matrix was developed by Dempsey and Baumhoff (1963). Their correlation coefficient is based on whether or not a type is found at a pair of sites; thus the name commonly used is Presence-Absence method. The advantage of employing this type of coefficient is that it does not give a bias in the matrix to types which are numerically important, but less so chronologically. In this technique, each type has equal weight. Hole and Shaw found that in one of their data sets, Pa Sanger flints, this procedure was more sensitive to non-lenticular variation (1967: 64, 78). However, it should be used only with caution. Methodologically, it is less misleading to work with the actual numerical occurrence than to tabulate presence/absence. In most cases, absence takes on too great a significance; a type may not be recovered on an archaeological site only because it is scarce (Cowgill 1968).

An accurate mechanical tool, easily used in the field, is the Meighan or Three–Pole system (1959). It incorporates percentages of the three main types in any population group. The percentages of these three types are computed as if they equaled 100% of the assemblages. These are plotted on triangular coordinate paper with a straight line drawn through the points which represent an approximate ordering. Meighan's system is well suited to preliminary analysis in the field because it is quick and requires no special equipment.

Underlying concepts of seriation

The assumption basic to seriation studies is that of stylistic replacement through time and space. A type begins its popularity at one site with few examples, reaches a maximum and slowly dies out. Another form reaches its zenith while the original type is losing its vogue. During the same developmental period, the type spreads in popularity to other sites at varying rates with the furthest sites receiving the original model last. The relative popularity of the class is usually shown by computing the percentage of each type at each site and plotting this on a graph. If replacement is taking place, this appears as a lenticular curve (for replacement of pottery wares in Korucutepe, see Kelly-Buccellati n.d.).

This model was tested in an interesting study analyzing dated gravestones in Colonial New England (Dethlefsen and Deetz 1966). The styles of these gravestones from a number of cemeteries were noted; their location and date plotted on a graph. The lenticular curve was produced, confirming assumptions about successive stylistic replacement in space, time and form (Dethlefsen and Deetz 1966: 504-5). The study demonstrated the diffusion rate of each type and underscored the problem of distinguishing between the effects of time and space on seriation. Both of these factors are influenced by a phenomenon known in physics as the Doppler effects (Deetz and Dethlefsen 1964; Clarke 1968: 426-7 and 462). The action of the Doppler effect can be seen only over a number of sites. The rate of change increases above the true rate as the initial site is approached from the most distant site; it decreases at a rate substantially below the true rate going away from the initial site.

Tests of the reliability of seriation techniques appear in an important study by Hole and Shaw on excavated material from Deh Luran (1967). They reasoned that if seriation was a reliable tool, it should arrange their data in the same order as it had been found in the stratigraphy, unless some other plausible explanation was forthcoming from the data itself. Their results showed that certain of the tests duplicated the stratigraphy, others only approximated it, and the remainder were not suitable. Ceramic and stone tool data from their sites were all seriated with the same five techniques, two of which they developed. Data which did not seriate (i.e. bad data) resulted from several factors: insufficient variation during the time involved, random occurrence at the site, a temporal change which did not follow the expected lenticular pattern (Hole and Shaw 1967: 36-7). Other factors to be considered include the possibility that specialized areas or 'activity areas' were sampled, thus reflecting a functional difference rather than a chronological one. Hole precludes this from his data by stating that all his samples came from midden areas (p. 6).

Implications for Ancient Near Eastern material

Dethlefsen and Deetz also show how useful seriation can be in assessing stylistic change (1966). In terms of Ancient Near Eastern material, this means that seriation may be extended to various quantifiable sets or classes of artifacts such as stone vessels or cylinder seals. The latter are particularly suited to seriation for the following reasons:

1) the body of material is large and thus, quantifiable;

2) the basic stylistic chronology is understood;

3) a large number from some periods can be dated by reference to their inscriptions or the dated tablets on which they have been rolled; and

4) their publication tends to be fuller than other classes of artifacts because glyptic holds interest for a variety of scholars.

First, the analysis of dated seals and impressions from one city should show their precise chronological position. A second step based on these results incorporates seals dated stylistically to the same period from that city or regions immediately surrounding it. Work is already underway in developing the technical codes needed for this analysis by the Centre d'Analyse Documentaire pour l'Archéologie (Gardin 1967). Encouragement is given by the excellent results of the Colonial gravestone study which go beyond chronological and stylistic change to shed light on political and social dynamics (Dethlefsen and Deetz 1966). Other research has begun with an analysis of the co-association of symbols and materials of Minoan seals to show regional variation (Reich and Morgan 1967, 1968, 1969).

Horizontal Patterning: Non-Random Distribution

Definition and historical background

A large body of literature is now available describing the theoretical basis and practical application of statistical techniques to give a more precise answer to a variety of questions (Sackett 1966; Clarke 1968; Binford and Binford 1968; Binford 1972). These questions are partly based on the view that the combination of artifacts and their spacial distribution reflect behavioral patterns, as well as general cultural patterns. All observable data are used to reconstruct ancient societies. In addition, the horizontal and vertical distributions of the data are as important as the data itself. Statistics is a necessary tool in the assimilation of this vast amount of information.

Some of the earliest applications of statistics in archaeological investigation are found in studies focusing on two widely separated areas: Predynastic Egypt (Meyers 1950) and prehistoric America. The impetus for the latter may have come from the need to organize and interpret the assemblages, artifacts and features from the pre-World War II government sponsored Works Progress Administration (WPA) projects in archaeology. Heretofore the implication had been that inferences about the past were limited because of the absence of adequate data (Smith 1955). Yet, new explanations about American pre-history were not generated by the wealth of detail resulting from the WPA projects. Perhaps stemming from this imbalance, archaeologists were urged to shift their focus from intersite comparisons to intra-site analysis (Taylor 1948). This meant a change from descriptions of artifacts from different sites to descriptions and explanations of the co-associations of artifacts and features from a single site. The design and testing of innovative approaches stressing problem-oriented research ('new archaeology'; e.g. Watson 1972) is a current development in the continuing re-evaluation of the theory behind and methods of archaeological research.

A good example of problem-oriented research is Hill's work at a pueblo of the 12-13th C., A.D. in the American Southwest. Some of his questions revolved around explanations of the formal variability of rooms (1968). A statistically representative sample of rooms was selected for excavation and study in order to ascertain the varying pattern of cultural remains.

On the basis of this evidence, Hill then offered predictions to explain variability and tested these inferences. His field research was designed to collect comparable data for statistical testing. The associations between and among the classes of artifacts and features were examined in order to isolate significant patterns of co-variation. For example, the following classes were tabulated: room size, floor area; presence/absence of fire pits, mealing bins (for grinding), doorways; style of masonry; sherd types per room, density of sherd per room, density per square meter on site; faunal and floral remains. Statistical evaluation of all of these paired and grouped co-associations was undertaken. Hill was able to explain the room variability in functional terms by referring to the attributes measuring these differences and to the tests of statistical inference. He further tested his explanations by reference to ethnographic evidence. His work yielded a wealth of information about past pueblo habitation.

Preliminary tabulation and description

Certain methods are basic once statistics are employed. Most researchers begin with the design of an attribute system, an organized code for tabulating and describing features or artifacts. Such a program forces the worker to examine the material, to record parameters carefully and explicitly state the criteria for each attribute. The attribute system presents what the scholar proposes as the limits of variability against which he measures the collection. There are no rules for selecting attributes; it is a matter of testing for rejection or inclusion, just as in a traditional trial sorting. But there is one overarching rule; the code must be mutually exclusive and mutually inclusive at the same time. Each set of attributes *must include* variables allowing for measurement of each and every member of the collection; but at the same time, an individual item may only be exclusively measurable by *one* of these attributes in a given set (Sackett 1966: 359-361; Elster 1971: 19-20). The attributes outlined in the system are those which are selected to best describe the collection in terms of the questions posed. For problems of prehistoric technology, one set of attributes applies; for stylistic variation, another set of variables is summarized. Once the attribute system is adopted, the data processing is explicit and objective with observations converted into numbers appearing as frequency distributions.

The techniques of statistical description generally utilize percentages of groups, industries, types, etc. based on frequency distributions. This data is summarized on bar histograms, cumulative frequency graphs or pie charts. Such aids visually present comparative differences and similarities.

Isolation of significant patterns: the Chi square test

Methods of statistical analysis may be employed if the relationship between the samples and the population from which they are derived is ambiguous. Such procedures select formulae which assist the archaeologist in evaluating and and analyzing the data. Basic to this is the Chi square test of significance, very useful in dealing with ambiguous associations (Spaulding 1953: 305-313; Sackett 1966: 365). For example, pottery types are observed in rooms and other loci in the excavation of a site. The question then arises: are we dealing with a random distribution or with a significant non-random pattern of association which should be investigated further? The reader may have noted Chi square referred to elsewhere as the 'null hypothesis' which only means that there is no relationship between, for example, the pottery types and the locations. If an association does exist, then such patterning will yield a frequency which is much greater than may be explained by chance alone. The Chi square test contrasts the observations reported by the archaeologist (the tabulated distribution) to that which chance alone dictates. With this statistical tool, the observed frequency is measured against the expected figure (due to chance alone) and this deviation becomes the Chi square score. Its significance is easily judged by reference to any standard Chi square table found in all basic statistics books (e. g. Blalock 1960: 452).

Every acceptable Chi square score is the result of an underlying pattern of co-variation. The tests, Cramer's V or Phi both assess the strength of this association (Sackett 1966: 367).

The raison d'être of the Chi square test is the hypothesis of chance. But since the hypothesis *is* chance or randomness, sampling error distorts the results if the total sample number is low. For example, with one hundred tosses of a coin, heads or tails probably appear in a 50:50 ratio. But with only 25 tosses of a coin, the ratio does not hold. Thus Chi square testing is not as reliable with small sample numbers. However, there is a formula defined as the Yates correction (Blalock 1966: 220-221), which is applied to serve as a control for small sample. In the Korucutepe examples given below, the Chi square test procedure and the application of Phi is outlined.

Technique of Chi square testing: Korucutepe examples

A combined team from the Universities of Chicago, California and Amsterdam joined during the seasons 1968-70 to investigate the mound at Korucutepe (van Loon 1969; van Loon and Buccellati 1970; van Loon and Güterbock 1970; van Loon 1971; van Loon and Güterbock 1972). The tell is sited in the Altinova plain near Elazig in the Keban Dam area, Turkey. Occupational debris represents settlement of the Chalcolithic, Bronze Age, Hittite and Seljuk periods. Our sampling area consists of the archaeological units reported as representing Early Bronze II and III. Radio-carbon determinants place the occupation in the second half of the third millennium (van Loon and Güterbock 1970: 126). The EB II and III areas excavated are spacially separated on Korucutepe (van Loon and Güterbock 1970: 3; van Loon 1971: 60, 61; van Loon and Güterbock 1972: 128).

Preservation and recovery is such that architectural features are seen to delineate specific areas. For the EB II, the team uncovered 4 rooms or houses surrounded by yards and 1 courtyard with features generally associated with food preparation. For the EB III there were 3 rooms or houses, 1 Shrine or hall joined to a corridor with a row of hearths and associated outside areas. These areas must reflect only a fraction of the occupation and are not to be compared to the kind of sample defined in Hill's investigation (1968). Furthermore, Hill's project carried a quantitative approach since he first isolated all of the rooms in the pueblo and then chose a statistically representative sample to excavate.

The Korucutepe areas were excavated under different conditions, but the pottery is quantifiable since all sherds were collected and careful tally kept of recovery from the numerous loci either from within a room (henceforth Inside) or from without (Outside). The sample consists of: type I — black burnished; II – red burnished; III – brown burnished; IV – painted; V – cream slipped; VI – unburnished, coarse; VII – imported; VIII – red/black burnished; IX – miscellaneous [1].

An example is given of the application of statistics to this quantitative data. Our question is whether the ceramic types are found in a patterned non-random distribution with the various Early Bronze II and III areas on Korucutepe. The pottery counts from each area are transferred to a frequency distribution table, which then becomes the data base for all subsequent computations. In Figure 1 the sample from all EB II units separated into Inside or Outside areas is tabulated.

EB II	I	II	III	IV	VI	VII	VIII	IX	Row Totals
INSIDE	321	99	416	14	45	14	130	208	1247
OUTSIDE	2692	265	1395	36	188	59	933	300	5868
(Column Totals)	3013	364	1811	50	233	73	1063	508	7115 (N)

EB III	I	II	III	IV	V	VI	VII	VIII	IX	Row Totals
INSIDE	164	32	253	18	165	212	3	101	25	973
OUTSIDE	103	45	149	5	54	188		130	5	679
(Column Totals)	267	77	402	23	219	400	3	231	30	1652 (N)

Fig. 1: Distribution Table of Pottery Types for EB II and EB III

A visual statistical description of these observations is reproduced in bar-histogram form in Fig. 2.

The question as to whether these frequencies represent random patterning or significant co-associations is taken up by means of the Chi square test. All of the computations may be performed manually on a desk calculator. In this case, use is made of the Olivetti Programma 101. The 101 is a mini-computer using the specific Programma language. The program for Chi square, Yates' correction and Phi (to measure the magnitude of association) was developed by Harold Kushner. The Olivetti 101 reduced many hours from the time required to compare ware types from area to area and determine the significance of co-association.

Fig. 1 exhibits the observed pattern of co-variation of all EB II Inside or Outside areas with ware types. In Fig. 3, comparison is made of what is *observed* (O column) against what would be *expected* (e column) on the basis of randomness or chance using the Chi square statistic; R stands for Row Total.

This contingency table shows the addition of the *expected* frequencies. Computation of the expected frequency (e) for each pair utilizes the formula:

[1] (The sample used in this study incorporates only those sherds found in association with rooms, thus the totals in this article differ from those in Kelly-Buccellati n. d.)

$$e = \frac{C \times R}{N}$$

(Blalock 1960: 215, 216). Here the border totals (from Fig. 1) of columns (C), e.g. 3013, and rows (R), e.g. 1247, are used while 'N' refers to total number in sample, in our case 7115. The observed frequency (the "O" column) and the expected frequency (the 'e' column) are contrasted in computing the Chi square score using the formula:

$$\chi^2 = \frac{(O - e)^2}{e}$$

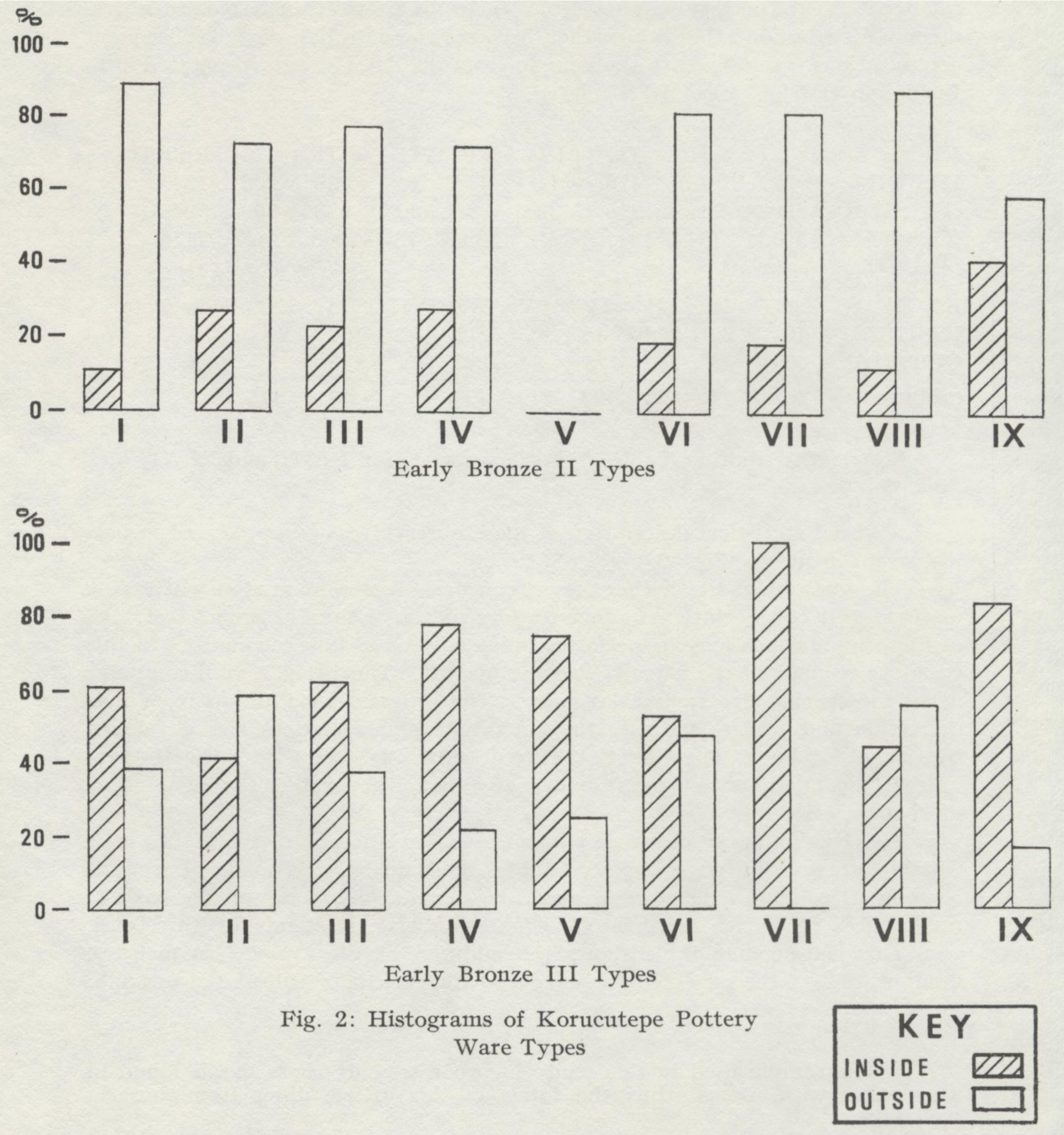

Fig. 2: Histograms of Korucutepe Pottery Ware Types

		O	C	R	e	χ^2
I	I	321	3013	1247	528.07	81.19
N	II	99	364	1247	63.79	19.43
S	III	416	1811	1247	317.40	30.62
I	IV	14	50	1247	8.76	3.13
D	VI	45	233	1247	40.83	.42
E	VII	14	73	1247	12.79	.11
	VIII	130	1063	1247	186.30	17.01
	IX	208	508	1247	89.03	158.97
O	I	2692	3013	5868	2484.93	17.25
U	II	265	364	5868	300.20	4.12
T	III	1395	1811	5868	1493.58	6.50
S	IV	36	50	5868	41.23	.66
I	VI	188	233	5868	192.16	.09
D	VII	59	73	5868	60.20	.02
E	VIII	933	1063	5868	876.69	3.61
	IX	300	508	5868	418.96	33.71

N = 7115 χ^2 = 376.90

Fig. 3: Contingency Table General Test I, EB II

(Sackett 1966: 367). Following the formula, the difference between observed frequency and expected frequency is squared, then divided by the expected frequency, giving the Chi square value for that pair. The sum of the individual Chi square scores for each contrasted pair results in the total Chi square value for the test (in our case, 376.90).

The significance of the total Chi square value of Test I, 376.90 is scored by reference to a standard table of Chi square values (e.g. Blalock 1960). At this point the 'degrees of freedom' (df) for the distribution table (Fig. 1) are required as an index to the Chi square score. The formula is: df = (C-1) × (R - 1), or degrees of freedom equals total number of columns (here 8) minus 1, times total number of rows (here 2) minus 1, the total here being 7 df. This indicates that if expected frequencies for 7 cells in the distribution table (Figure 1) are known, the 9 remaining may be established by substraction. This rule applies to any size distribution table.

Test I, for the EB II, scores at the .001 'confidence level' which means there is but one chance in a thousand that the association of ware types and locations is due to pure chance. In other words, we have been questioning whether the distribution of pottery is random or non-random and with a level at .001 it can be confidently stated that non-randomness has been demonstrated.

If the score were significant at the .01 confidence level, there would be 10 chances in a thousand (or one in a hundred) that this distribution is due to chance alone. The confidence level considered significant is an arbitrary decision (Sackett 1966: 376); in this study both the general and restricted tests have a confidence level of over one percent.

Examination of Fig. 3 indicates which types and associations contribute most to the Chi square score. These are summarized in descending order of their contribution to the total Chi square score in Fig. 4. Pairs below the dotted line represent observed and expected frequencies so close as to be virtually random.

If observed frequency is much less than expected, a negative-inflated Chi square score may be the result (see Fig. 3, type I/Inside). Such pairs are not referred to in our summary, although their scores add to the total. As a control, inflated scores are added together, then subtracted from the Chi square total. This new total is compared to a Chi square table of values to see whether the confidence level is acceptable and therefore the test results. Of course the absence or low frequency of certain types may have archaeological significance and should be examined.

The continuity requirements for Chi square testing are (1) none of the expected frequencies may fall below 1.0; and (2) no more than 20% of the expected frequencies may fall below 5.0 (Sackett 1966: 369). If the requirements are not fulfilled, the test may not be reliable. However, if examination shows no inflation (of total Chi square score), results are probably acceptable. As the worker experiments with the Chi square statistic, many of these points fall into place.

χ^2 .001 ϕ .23	χ^2 .001 ϕ .21
IX : Inside	VIII: Outside
III : Inside	V : Inside
II : Inside	II : Outside
I : Outside	VI : Outside
VIII: Outside	IX : Inside
IV : Inside	IV : Inside
. . . .	III : Inside
VI : Inside (very low score)	. . . .
VII : Inside (very low score)	VII : Inside (very low score)
V : Not present	I : Inside (very low score)
General Test I	General Test II
Early Bronze II	Early Bronze III

Fig. 4: Rank Order of Chi Square Scores and Significant Associations

Phi measures the association between variables such as ware types and locations, but the Phi value is not the measure of the strength of just one pair (for example, type II: Inside). Rather, Phi assesses all the underlying paired co-associations which contribute to the Chi square score. The formula is:

$$\phi = \sqrt{\frac{\chi^2}{N}}$$

Scores are scaled from 0 to 1. A series of Phi scores is useful for comparing the strenght of association from a group of tests. In addition, Phi controls for sample size because, in the formula, the Chi square score is divided by the number (N) in the sample; thus Phi scores may be compared although they derive from Chi square tests on populations of various sizes. Chi square, on the other hand, is sensitive to sample number and may yield higher Chi square scores with larger populations.

Contrast between descriptive and analytic statistics

A graphic example of the difference between descriptive statistics and analytic statistics appears with the comparison of Fig. 2 to Fig. 4. Both deal with the same data but the bar histograms of Fig. 2 describe precisely what percentage of each pottery type is recovered from Inside or Outside locations. The greater recovery of a ware type from a particular location does not necessarily predict a significant co-association. For example, types II, III, IV, IX hold higher percentages from EB II Outside locations (Fig. 2); yet the statistically significant non-random association of these types is with Inside areas (see Fig. 4). A parallel situation can be seen for the EB III with type VI and the Inside. Fig. 4 summarizes the significant co-associations of types and locations (in rank order) as isolated in the Chi square test in Fig. 3. The order of these co-associations could not have been inferred from the descriptive bar histograms of Fig. 2.

The patterns demonstrated are clearly seen on the following chart, Fig. 5, rearranged from Fig. 4 to point up the parallels. The figure in parenthesis to the left of the type numeral refers to the total percentage of this type in the population (from *all* loci).

EB II	EB III
Inside	Inside
(.26) III	(.24) III
(.07) IX	(.02) IX
(.001) IV	(.01) IV
(.05) II	(.13) V
Outside	Outside
(.42) I	(.14) VIII
(.15) VIII	(.05) II
	(.24) VI
Random	Random
(.03) VI	(.001) VII
(.01) VII	(.16) I
Not Present	
V	

Fig. 5: Inside-Outside Type Associations for EB II and III

To summarize from EB II to III:

1. No change is demonstrated in location and very little in percentage for types III, IV, VIII, IX.

2. The standard EB II household wares, III, IV, and IX continue into the EB III period. Likewise type VIII maintains its relative percentage and association with Outside areas through time.

3. Type V enters the inventory of houses with EB III. This development is explained as a function of time.

4. Changes seen for both common and uncommon types:

a) Type I loses in percentage through time (.42 in EB II; .16 in EB III) and changes association from Outside to Random.

b) Type II (.05) changes association from Inside to Outside but maintains degree of percentage in both populations.

c) Type VI (.03), randomly distributed in EB II, increases eight fold (.24) in the EB III population and is found in association with Outside areas.

Restricted Chi square tests

With these associations demonstrated, two restricted tests are presented (for EB II and III) which compare the incidence of ware types from specific Inside and Outside locations. The expectation is that the associations isolated in the general tests will duplicate in the specific tests comparing a single room or a few rooms to a related outside area. Recovery from EB II Rooms 1, 2, 3, 5 is compared with that from the Courtyard. For the EB III, wares collected from Room 15 are contrasted with those from the area outside. Test summaries are seen in Fig. 6; the calculations are omitted.

Test III, EB II		Test IV, EB III	
χ^2 .001 ϕ .32		χ^2 .001 ϕ .37	
Rms 1, 2, 3, 5 to Courtyard		Room 15 to area Outside	
I	Courtyard	II	Outside
IX	Rooms	VI	Outside
III	Rooms	V	Room
II	Rooms	III	Room
VII	Rooms	IX	Room
VI	Rooms (very low score)	I	Outside (very low score)
IV	Rooms (very low score)	VIII	Outside (very low score)
		VII	Room (very low score)

Fig. 6: Restricted Tests III and IV in Rank Order of Chi Square Score and Significant Associations (Rooms 1-3,5 are in squares 012, N11, N12, 016 and 017; the courtyard is in N11, 12; Room 15 is in O14 and P14).

Comparison of Fig. 6 to Fig. 4 shows fair duplication of General Tests with the following points to note:

1. In Test III type VII has a significant co-association with the Rooms, whereas in Test I the Chi square score for this pair is very low. Thus Test III adds needed support to this type-association.

2. The anomaly in Test IV is Type I which again scores very low but this time with Outside. Since Type I scores low on both Test II and IV we are inclined to see the random distribution of this type as a function of sample size.

To summarize: the distribution of certain pottery types in room-houses and outside areas of Early Bronze II and III Korucutepe is non-random. The ambiguous pattern has been clarified using a method of statistical analysis. The question of patterning cannot be completely explained but is carried as far as the data allows.

Conclusions

All archaeologists are concerned with the co-associations exposed through survey and excavation. Here we have tried to show how statistical techniques can clarify ambiguous patterning: both vertical and horizontal. The inferences thus generated are not necessarily different from those reached without this tool. In fact, statistics does not always show something new nor does it use different logic. Still, an inference thus derived has an added dimension which is not negligible: it is the addition of a precise and explicit quality to those inferences.

The insistence upon explicitness and explanation is underscored as the discipline of archaeology broadens to include the aims and methodologies of the natural sciences (Adams 1968; Hammond 1971; Watson, LeBlanc and Redman 1971). Not all archaeologists will agree with this diversification in the process and practice of archaeology (Hawkes 1968). But a report on the quest for an expanded relevance for archaeology is germane to the central concern of this volume. Scholars will want to examine these expanded goals and their bearing upon the manner and means of excavating in the Ancient Near East.

Bibliography

The authors wish to thank Dr. James R. Sackett for his many invaluable suggestions. We also wish to express our appreciation to Mr. Sandy Elster for extending stenographic assistance and access to the Olivetti Programma. Readers interested in receiving the Program are encouraged to correspond with the authors.

Adams, Robert McC. (1968) "Archeological Research Strategies: Past and Present", *Science* 160, No. 3833: 1187-92.

Ascher, Marcia and Robert Ascher (1963) "Chronological Ordering by Computer", *American Anthropologist* 65: 1045-52.

Binford, Lewis R. (1972) *An Archaeological Perspective.* New York.

Binford, Sally R. and Lewis R. Binford, eds. (1968) *New Perspectives in Archeology.* Chicago.

Blalock, H. M. Jr. (1960) *Social Statistics.* New York.

Brainerd, G. W. (1951) "The Place of Chronological Ordering in Archaeological Analysis", *American Antiquity* 16: 301-13.

Clarke, David L. (1968) *Analytical Archaeology.* London.

Cowgill, George L. (1968) review of Frank Hole and Mary Shaw, *Computer Analysis of Chronological Seriation*, *American Antiquity* 33: 517-9.

Deetz, James (1967), *Invitation to Archaeology*, New York.

Deetz, James and E. Dethlefsen (1965) "The Doppler Effect and Archaeology: A Consideration of the Spacial Aspects of Seriation", *Southwestern Journal of Anthropology* 21: 196-206.

Dempsey, Paul and Martin Baumhoff (1963) "The Statistical Use of Artifact Distributions to Establish Chronological Sequence", *American Antiquity* 28:496-509.

Dethlefsen, E. and James Deetz (1966) "Death's Heads, Cherubs and Willow Trees: Experimental Archaeology in Colonial Cemeteries", *American Antiquity* 31: 502-10.

Dyson, Robert H. (1968) "The Archaeological Evidence of the Second Millennium B. C. on the Persian Plateau" *C.A.H.* [2], Fascicle 66. Vol. II, Ch. XVI.

Elster, Ernestine S. (1971) "Sitagroi 'Cult' Vessels. A Comparative Typological Study Using Attribute Analysis", U.C.L.A. Thesis, unpub.

Gardin, Jean Claude (1967) "Methods for the Descriptive Analysis of Archaeological Material", *American Antiquity* 32: 13-30.

Gardin, Jean Claude, ed. (1970) *Archéologie et Calculateurs: Problèmes Sémeologiques et Mathématiques*, Paris.

Gimbutas, Marija (1970) "Proto-Indo-European Culture: The Kurgan Culture during the Fifth, Fourth and Third Millennia", in G. Cardona, H. M. Hoenigswald, and A. Senn, eds., *Indo-European and Indo-Europeans*, Philadelphia 155-197.

Hammond, Allen (1971) "The New Archaeology: Toward a Social Science", *Science* 172, No. 3988: 1119-20.

Hawkes, Jacquetta (1968) "The Proper Study of Mankind", *Antiquity* XLII: 255-262.

Hill, James N. (1968) "Broken K. Pueblo: Patterns of Form and Function", in S. Binford and L. Binford, eds. *New Perspectives in Archeology*, Chicago, 103-42.

Hodson, F. R., D. G. Kendall and P. Tãutu, eds. (1971) *Mathematics in the Archaeological and Historical Sciences*, Edinburgh.

Hole, Frank and Mary Shaw (1967) *Computer Analysis of Chronological Seriation*, Rice University Studies 53, No. 3, Monograph in Archaeology, Houston.

Kantor, Helene J. (1965) "The Relative Chronology of Egypt and Its Foreign Correlations before the Late Bronze Age", in Robert W. Ehrich, ed. *Chronologies in Old World Archaeology*, Chicago, 1-46.

Kelly-Buccellati, Marilyn (n.d.) "Early Bronze Pottery from the Three Seasons 1968-1970 at Korucutepe", *JNES*.

Kendall, David G. (1963) "A Statistical Approach to Flinders Petrie's Sequence Dating", *Bulletin of the International Statistical Institute: Proceedings of the 34th Session*, XL, Pt. 2: 657-81.

Kroeber, A. L. (1916) "Zuñi Potsherds", *Anthropological Papers of the American Museum of Natural History*, 18, Pt. 1: 1-37.

Kuzara, R. S., G. R. Mead, and K. A. Dixon (1966) "Seriation of Anthropological Data: A Computer Program for Matrix–Ordering", *American Anthropologist* 68: 1442-55.

Massoulard, Emile (1949) *Préhistoire et Protohistoire d'Egypt*, Université de Paris: Travaux et Mémoires de l'Institut d'Ethnologie, LIII, Paris, especially 61-69.

Meighan, C. W. (1959) "A New Method for the Seriation of Archaeological Collections", *American Antiquity* 25: 203-211.

Meyers, Oliver H. (1950) *Some Applications of Statistics to Archaeology*, Service des Antiquités de l'Egypte, Cairo.

Petrie, W. M. Flinders (1901) *Diospolis Parva*, Egyption Exploration Fund, London, especially 4-12.

Petrie, W. M. Flinders (1899) "Sequences in Prehistoric Remains", *Journal of the Anthropological Institute of Great Britain and Ireland*, 29 (N. S. II) 295-301.

Porada, Edith (1970) *Tchoga Zanbil IV: La Glyptique*, Mémoires de la Délégation Archéologique en Iran, Vol. XLII (Mission de Susiane), Paris.

Reich, J. and R. Morgan *CALCULI*, March 1967: 5; July 1968: 48; Sept. 1968: 56; March 1969: 68.

Robinson, W. S. (1951) "A Method for Chronologically Ordering Archaeological Deposits", *American Antiquity* 16: 293-301.

Sackett, James R. (1966) "Quantitative Analysis of Upper Paleolithic Stone Tools", in J. Desmond Clark and F. Clark Howell, eds. *Recent studies in paleoanthropology*, American Anthropologist Special Publication, Vol. 68, No. 2, Pt. 2: 356-94.

Smith, M. A. (1955) "The Limitations of Inference in Archeology", *Archaeology News Letter* 6: 3-7.

Spaulding, Albert C. (1953) "Statistical Techniques for the Discovery of Artifact Types", *American Antiquity* 18: 305-13.

Spier, Leslie (1917) "An Outline for the Chronology of Zuñi Sites", *Anthropogical Papers of the American Museum of Natural History* 18, Pt. 3: 209-331.

Sterud, Gene (1967) "Seriation Techniques in Archaeology", U.C.L.A. Thesis, unpub.

Taylor, Walter W. (1948) *A Study of Archaeology*, American Anthropological Association Memoir No. 69, Reprinted 1967, Carbondale, Ill.

van Loon, Maurits (1971) "Korucutepe Excavations, 1969", *Keban Project 1969 Activities*, METU Keban Project Publications, Series 1, No. 2, Ankara.

van Loon, Maurits and Giorgio Buccellati (1969) "The 1968 Excavations at Korucutepe Near Elazig", *Türk Arkeoloji Dergisi* XVII-1: 79-82.

van Loon, Maurits and Giorgio Buccellati (1970) "The University of Chicago–University of California Excavations at Korucutepe – 1968," *1968 Summer Work*, METU Keban Project Publications, Series 1, No. 1, Ankara.

van Loon, Maurits and Hans G. Güterbock (1970) "The 1969 Excavations at Korucutepe Near Elazig", *Türk Arkeoloji Dergesi XVIII*-2: 123-28.

van Loon, Maurits and Hans G. Güterbock (1972) "The 1970 Excavation at Korucutepe Near Elazig", *Türk Arkeoloji Dergisi* XIX-1: 127-32.

Watson, Patty Jo, Steven A. Le Blanc and Charles L. Redman (1971) *Explanation in Archaeology*, New York.

Watson, Richard (1972) "The 'New Archeology' of the 1960s", *Antiquity* XLVI, No. 183: 210-15.

Whallon, Robert (1972) "The Computer in Archaeology: A Critical Survey", *Computers and the Humanities* 7: 29-45.