

## Making Sense of Contingency Tables in Archaeology: the Aid of Correspondence Analysis to Intra-Site Activity Areas Research

Gianmarco Alberti  
*PhD-University of Udine*

*Abstract:* The use of contingency tables is widespread in archaeology. Cross-tabulations are used in many different studies as a useful tool to synthetically report data, and are also useful when analyst wishes to seek for latent data structures. The latter case is when Correspondence Analysis (CA) comes into play. By graphically displaying the dependence between rows and columns, CA enables the analyst to explore the data in search of a meaningful inner structure. The article aims to show the utility of CA in archaeology in general and, in particular, for the identification of areas devoted to different activities within settlements. The application of CA to the data from a prehistoric village in north-eastern Sicily (P. Milazzese at Panarea, Aeolian Islands-Italy), taken as a case study, allows to show how CA succeeds in pinpointing different activity areas and in providing grounds to open new avenues of inquiry into other aspects of the archaeological documentation.

*Key words:* Correspondence analysis, contingency tables, archaeology, activity areas research, P. Milazzese, Panarea, Sicily, Middle Bronze Age.

### 1. Introduction

Archaeological data are inescapably numerical in nature. As stressed by Van Pool and Leonard (2011, p. 5), archaeologists measure and count everything, from pot sherds to bone fragments, from site features to whole buildings, with all that lies in between. It comes with no surprise if the need of statistical tools to make sense of data has steadily increased during the development of the discipline. Besides, archaeologists often happen to deal with categorical variables in that it is quite naturally for them to classify into categories the material reality they study (e.g., animal species, stone tool types, pottery types, vessel functions, etc.). It is not by chance if the contingency table is one of the most widespread means to summarize and present the data in archaeology.

Today, the analysis of contingency tables with a host of statistical techniques makes up lengthy sections of books devoted to the use of statistics in archaeology

(Baxter, 1994; Shennan, 1997; Drennan, 2009; Van Pool and Leonard, 2011). Beyond the use of hypothesis tests (namely, the chi-square test) to formally assess the degree to which rows and columns of the table are independent (first introduced in archaeology by Spaulding, 1953), during the last 30 years archaeology has witnessed the rise of interest in approaches capable to get the most of the contingency tables in terms of exploration of data structure. In particular, Correspondence Analysis (hereafter, CA) has proved to be a valuable tool for the interpretation of the complex datasets. The possibility to display rows and columns of a contingency table in graphical form enables the analyst to reduce the dimensionality of the data and to explore different trend of variability, allowing hidden patterns to emerge.

The application of CA has steadily increased in the social science (see, e.g., the various articles published in Blasius and Greenacre, 1998) as well as in archaeology. Even though in the latter field CA has been slow in gaining popularity, with the exception of early groundbreaking studies from continental Europe (Bølviken *et al.*, 1982; Djindjian, 1985; Madsen, 1989), today the CA is used for many purposes, ranging from burial assemblages analysis (Wallin, 2010), on-site distribution of faunal remains (Potter, 2000), distribution of pottery types in different kind of archaeological contexts (Cool and Baxter, 2002; Pitts, 2005), stratigraphy and formation processes (Mameli *et al.*, 2002; Pavùk, 2010), seriation and chronology (Kjeld Jensen and Høilund Nielsen, 1997; Smith and Neiman, 2007; Bellanger *et al.*, 2008; Peeples and Schachner, 2012).

## 2. Aim of the Article

The aim of the article is to describe the utility of CA in archaeology in general, and in the context of intra-site activity area research in particular. The latter refers to the study of the spatial distribution of the material remains of past activities, with the aim to backtrack the cultural processes that generated them (overview in Kent, 1987; Verhoeven, 1999, pp. 11-13; Steadman, 1996; Cutting, 2006, pp. 228-230).

Taking the prehistoric Middle Bronze Age settlement on the island of Panarea (Aeolian Archipelago, Italy) as a case study (stemming from a broader analysis of the settlements of this period in north-eastern Sicily made in Alberti, 2012), it will be shown how CA can be used to explore the relation between finding spots (i.e., huts) and functional classes of objects, in order to pinpoint the activities performed in the village's huts (or in groups thereof). It will be also shown how CA hallows patterns to emerge that turn up to be useful in the context of the interpretation of settlement activities and spatial organization, providing the basis for further speculations in the domain of the relationship between spatial and social organization of a past community. These latter aspects will be only

touched upon, since an account of the broader social implication of the on-site activity patterns is beyond the scope of this article.

In what follows, first, I will sketch up a brief jargon-free introduction to CA. It is not my intention to provide a full account of its theoretical and computational underpinnings, since these can be easily found in Greenacre, 2007 and, from an archaeological standpoint, in Baxter (1994, pp. 100-139) and Shennan (1997, pp. 308-341). Moreover, a number of articles in this same Journal have elegantly sketched up a description of the technique (Panagiotakos and Pitsavos, 2004; Blanco Abellán, 2007). After, the discussion of a case study will bring us into the core of the article's argument, allowing to present the advantages of CA in the activity areas research. I will first provide essential information on the site under study and on the data object of the analysis. The results of the CA will be then described, and the aid it provides in discovering latent patterns of association between huts and functional classes of objects will be discussed. It will be also shown how the insights stemming from CA can provide grounds to address further questions of archaeological interest like: (a) the existence of a difference in floor area between huts devoted to different purposes (i.e., dwelling vs. utilitarian cabins); (b) the estimate of the number of individuals inhabiting the settlement. Finally, conclusions will follow.

### 3. A Short Introduction to CA

CA is an exploratory technique aimed to graphically represent the dependence between rows and columns of contingency tables. The visual display of data helps the interpretation and allows patterns to emerge. The technique reduces the number of dimensions needed to display the data points by decomposing the total inertia (i.e., the variability) of the table and defining a smallest number of dimensions capable to capture the data variability. The graphical output of CA is a scatterplot where rows and/or columns are represented as points on a sequence of low-dimensional spaces. These spaces have the properties to retain a decreasing amount of the total inertia. The first dimension will capture the highest amount, while the second will be associated to the second largest proportion, and so on.

On the scatterplot, the distance between data points of the same type (i.e., row-to-row) is related to the degree to which the rows have similar profiles (i.e., relative frequencies of column categories). The same applies for the column-to-column distance. The more the points are close to one another, the more similar their profiles will be. The origin of the axes represents the centroid (i.e., the average profile), and can be conceptualized as the "place" where there is no difference between profiles or, more formally (and to recall the chi-square terminology), it represents the hypothesis of homogeneity of the profiles (Greenacre, 2007, p. 32). The more different are the latter, the more the profile points will be spread on

the plane away from the centroid.

As for the relative distances between points of different type (i.e., row-to-column), it tells the analyst something about the “correspondence” between the categories that made up the table. In other words, the more a row point is close to a column point, the greater (i.e., the more distant from the average) is the proportion that that column category makes up on the row profile.

#### **4. The Analysis of Contingency Tables and the Search of Activity Areas**

The reconstruction of forms of social organization at the on-site level hinges on the possibility to pinpoint the places where different kinds of activities were performed in the past, and to backtrack from these the forms in which ancient societies organized their tasks and activities, and, ultimately, their life, identity, and social relations. It becomes essential for the sake of any subsequent archaeological interpretation to understand what activities were performed in which locations, and by means of which objects or groups thereof. Acknowledging the fact that any inference must be preceded by the identification and understanding of the cultural and non-cultural process likely to have affected the original artefact inventories (Schiffer, 1989; Verhoeven, 1999, pp. 47-60), the prerequisite of any activity areas study is that the objects found in the huts’ usage level undergo a functional classification. In other words, objects are classified according to the purpose they were used to (Adams and Adams, 1991, pp. 221-223; Lowell, 1991, p. 20; Verhoeven, 1999, pp. 71-103).

This is exactly where the use of CA comes into play. In fact, in the study of the distribution of functional classes of object across village’s huts, the analyst is finally in the position to build a contingency table where the frequency of objects (used for different functions) across the finding spots (i.e., huts’ usage levels) is tabulated. It is then important: (a) to assess the strength of association (if any) between rows and columns of the contingency table; (b) for the sake of the identification of specific activity spots, to explore the “correspondence” between row and column categories, i.e., between huts and different types of objects. On this respect, while different technique has been developed in archaeology to seek for significant clusters in the horizontal distribution of objects across unbounded spaces on the basis of density figures (overview in Blankholm, 1991, pp. 169-178), CA turns out to be particularly well suited to the above situation since the huts (with their daily assemblages) are already discrete groups liable to be compared in search of similarity or difference in the proportion of different types of objects.

#### **5. CA and Intra-Site Activity Areas Research: A Case Study**

### 5.1 The Middle Bronze Age Settlement at P. Milazzese of Panarea

The settlement object of this article's analysis lies on the promontory named Punta Milazzese, on the island of Panarea (Aeolian Archipelago, Italy) (Figure 1A-B). It was unearthed by Bernabò Brea and Cavalier (1968) who led the excavations at the site bringing to the light the objects that made up the material inventories used by past people on a daily basis.

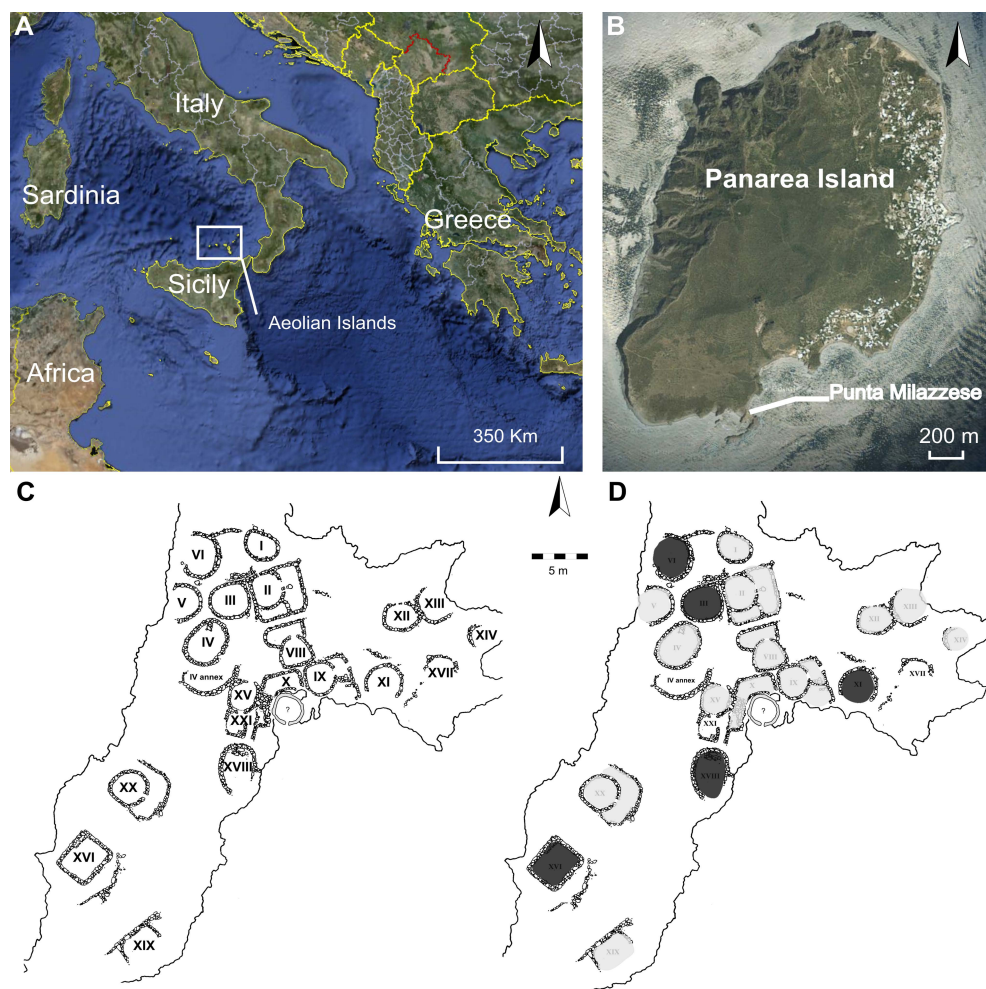


Figure 1: A) Central Mediterranean basin showing the location of the Aeolian Islands to which the settlement of P. Milazzese (Panarea) belongs. B) Panarea Island with location of the prehistoric settlement. C) P. Milazzese settlement plan (roman numbers indicating the huts according to Bernabò Brea and Cavalier, 1968). D) Settlement plan with huts coloured on the basis of the functional classification deriving from CA (dark grey: dwelling huts; light grey: utilitarian huts) (C-D after Bernabò Brea and Cavalier, 1968, modified)

The inventories were made up of clay objects as well as of stone tools. Ceramic vessels were very common, comprising pots of both local and non-local production. The latter group was made up of vessels coming from mainland Italy (Apennine pottery) and continental Greece (Aegean pottery).

The village dates to the local Middle Bronze Age (about 1460-1270 BC; Alberti, 2011, 2013) and is made up of about 24 huts built with the use of local raw materials (stone and pebbles for the walls, wooden and other perishable materials for the ceiling) (Figure 1C). The cabins are featured by different plans (ranging from round/oval to rectangular) and can often have annexes, which are likely to have been unroofed areas attached to the main room (Holloway and Lukesh, 1995, pp. 64-65).

## 5.2 In Search of Activity Areas: the Aid of CA

While the excavation report provided important information about the artefact inventories found in huts, in subsequent studies no attempt has been done to pinpoint specific activity areas within the settlement in order to understand: (a) if the activities were or not evenly distributed across the village; (b) if it is possible to identify differences in the huts' functions; (c) if it is possible to identify huts used mainly for habitation purposes as opposed to more utilitarian ones; (d) what relation (if any) existed between function and huts' dimension (a full account of the theoretical issues and a review of the literatures in Alberti, 2012). To address these questions, a specific study aimed to explore pattern of associations between huts and functions was needed. In what follows, it will be shown how CA can provide important insights into the above issues.

On the basis of the excavation report and of a close scrutiny of the layers where the objects were found, it has been possible to build up a  $31 \times 19$  contingency table (Table 1) where the frequency of objects with known function (rows) is tabulated against the finding contexts (i.e., huts, put in columns).

CA was performed on that table. It has to be noted that two type of data has been entered as supplementary points and does not affect the results of the analysis (Greenacre, 2007, pp. 89-96): (1) objects with dubious function (Table 1, rows 4-5, 15, 28-31); (2) huts for which there is reason to believe that their inventory was affected by post-depositional events (Table 1, last seven huts to the right). While, as stressed, those categories will not affect the CA results, the possibility to project them on the CA map allows to understand how they relate to the other categories displayed.

Table 2 reports the total inertia and the proportion of it accounted by the CA dimensions. Following Greenacre (2007, p. 28, p. 61), the square root of the total inertia can be considered a measure of the strength of association between rows and columns; in our case, the value is 0.867. It can be interpreted as pointing to

Table 1: Frequency of objects with different functions across the huts of the P. Milazzese settlement. The prefix A indicates objects from mainland Italy (i.e., Apennine culture); the prefix Ae indicates objects from Late Bronze Age Greece (i.e., Aegean culture)

object function	object	Huts																			Total	
		M01	M02	M03	M04	M05	M06	M08	M09	M10	M16	M18	M20	M04annex	M11	M12	M13	M14	M15	M19		
1	cooking	tray	2	5	3		1		2	4	2	1	1	2	1	3			1	1	29	
2	cooking	cooking vessel									1		1	3							5	
3	cooking	cooking stand		1				1													2	
4	cooking	Ae-tray								1											1	
5	cooking/suspending?	clay hook			2	1			1		1	1	3		3		1				10	
6	dinner	stand		1	4	2	2	1	1			3	2	1		2				1	1	21
7	dinner (drinking/eating)	A-open vessel		1	2	1	1		1	1	4	3	2	2		6		3	1	1	29	
8	dinner (drinking/eating)	Ae-open vessel									1	1	1			2					5	
9	dinner (eating)	fine-ware pedest. bowl	1	3	1	1	1	2	2	5	4	2	1	3		2		2		2	1	33
10	dinner (eating)	coarse-ware pedest. bowl								2	1		1	2	1				2		1	10
11	dinner (pouring)	A-closed vessel	1	1	2		1		1	2			1								9	
12	dinner (pouring)	fine-ware jug	1	3	4	3	2	2	2	4	4	2	5	5	1	7				1	1	47
13	dinner/processing	big eating/processing bowl	1			2	1		2					1						1	1	8
14	dinner/processing?	small pedest. bowl	1	3	2	1			2	1	2	1		1	1	1	1	1		1	1	19
15	dinner/processing	small bowl	1	3									1								5	
16	pouring	coarse-ware jug							1		1		1								3	
17	processing	mortar/pestel		2	2	2			1	12	3	1		2	1	2				1	29	
18	processing	millstone/handstone		4	1				1	3	1			6							16	
19	spinning	spindle worls								2	5								2		1	10
20	storing	big stor. vessel		2	2	1		1	2	2	2	4		1	1	1				1	1	21
21	storing	Ae-closed vessel		1	1					1	1	1	1		1						1	8
22	storing	small stor. vessel	2	2	1	2	1	1	3	4	2	1	2	3		1		1		1	1	27
23	storing/cooking	small stor./cooking vessel	1								2										3	
24	storing/processing	small stor./process. vessel	2	2	1	1		1		1	1		1								1	11
25	storing and other func.	stor. and other funct. vessel	2	3	1	1	1		2	2	1	1	1		1		1	1	1	1	1	20
26	working	stone tool		2	1		1		2	3	4											13
27	working	lithic core/flake		4	7		1	6	1	7	8	5	5	6					3			53
28	covering	A-lid				1															1	
29	working?	clay lid/disc/sieve			1	2	1			1											5	
30	working/playing/ritual?	miniature vessel	1			3				1	1		3		4		1	1		1	18	
31	working?	no-worn pebble						12	1						5						18	
<i>Total</i>			16	43	39	23	14	28	25	60	53	29	23	44	11	37	3	14	5	9	13	489

a strong association, judging on the basis of a five-tiered scale (weak: 0.0-0.20; low: 0.20-0.40; moderate: 0.40-0.70; strong: 0.70-0.90; very strong: 0.90-1.00) (Rowntree, 2000, p. 170; for other scales and for the problem of the interpretation of measures of association, see e.g., Reynolds, 1977, pp. 30-34; Taylor, 1990, p. 37).

As for data interpretation and the choice of the number of dimensions relevant to it, if there were no association between rows and columns, each dimension

Table 2: Inertia accounted for by the CA dimensions. Proportion and cumulative proportion of the inertia accounted for by each dimension are also shown

Dimension	Inertia	Proportion (%)	Cumulative (%)
F1	0.153	20.38	20.38
F2	0.141	18.68	39.06
F3	0.121	16.11	55.17
F4	0.095	12.63	67.80
F5	0.073	9.65	77.45
F6	0.051	6.73	84.18
F7	0.037	4.86	89.04
F8	0.034	4.49	93.53
F9	0.029	3.90	97.43
F10	0.012	1.54	98.97
F11	0.008	1.03	100.00
Total	0.753		

would explain the same proportion of inertia. In our example, each axis would account for  $1/(24-1) = 0.043$  of the inertia in terms of (active) rows,  $1/(12-1) = 0.090$  in term of (active) columns. Any axis contributing more than the higher of those two figures should be considered important for the interpretation of the data (Bendixen, 1995, p. 577). The first four dimensions (together accounting for the 67.80% of the inertia) can be considered important for the interpretation of the table's structure. It must be acknowledged, however, that the number of dimensions to keep is fixed to the very analyst's ability to give meaningful interpretation of the axes kept for the analysis, as stressed by various scholars (Benzécri, 1992, p. 398; Clausen, 1998, p. 25; Yelland, 2010, p. 13). It will be evident later on how we can arrive to a meaningful (in archaeological terms) interpretation by mainly focusing on the first two dimensions.

As for dimensions interpretation, since the interest here lies in understanding the degree to which huts are different in the proportion of different objects, I decided to interpret the scatter of column points (i.e., huts) in the space defined by the row categories (i.e., objects).

Table 3 reports the row categories (i.e., objects) having a higher-than-average contribution (Greenacre, 2007, p. 82) to the definition of the first four dimensions. These figures can be taken into account along with the CA symmetric map in Figure 2, showing the column points in the space defined by the first two dimensions.

The first dimension is determined by the opposition between objects linked to function like small storing, storing/cooking, processing, and working (stone tools) on the one hand (positive pole), and objects linked to consumption practices (fine-ware drinking/eating vessels of non-local origin) and production of stone objects,



Table 3: Contributions of the row categories (i.e., archaeological artifacts) to the definition of the dimensions relevant to data interpretation. In bold: higher-than-average contributors (i.e., categories having major contribution to the definition of that particular dimension)

Row Contributions to the Inertia (average row contribution = 0.042)				
	F1	F2	F3	F4
A-open vessel	<b>0.046</b>	0.000	0.029	0.001
tray	0.034	0.008	0.010	<b>0.050</b>
mortar/pestel	<b>0.072</b>	<b>0.066</b>	0.014	<b>0.207</b>
stone tool	<b>0.049</b>	0.015	0.036	0.028
big stor. vessel	0.025	0.003	0.000	0.022
stand	<b>0.138</b>	<b>0.160</b>	0.000	0.007
stor. and other funct. vessel	<b>0.045</b>	<b>0.093</b>	0.000	0.018
small pedest. bowl	0.021	0.016	0.001	<b>0.075</b>
millstone/handstone	0.007	<b>0.080</b>	<b>0.254</b>	0.031
small stor./process. vessel	<b>0.044</b>	0.038	0.001	0.000
A-closed vessel	0.014	0.002	0.010	0.001
big eating/processing bowl	<b>0.073</b>	<b>0.121</b>	0.000	<b>0.283</b>
Ae-closed vessel	0.016	0.000	0.013	0.001
Ae-open vessel	<b>0.056</b>	0.001	<b>0.058</b>	0.000
small stor./cooking vessel	<b>0.061</b>	0.001	<b>0.137</b>	0.006
fine-ware jug	0.036	0.016	0.001	0.019
fine-ware pedest. bowl	0.001	0.004	0.000	0.000
coarse-ware pedest. bowl	0.001	<b>0.074</b>	0.014	0.034
small stor. vessel	0.012	0.020	0.010	0.003
spindle worls	0.037	<b>0.161</b>	<b>0.242</b>	0.003
lithic core/flake	<b>0.177</b>	0.032	0.015	0.000
cooking vessel	0.019	<b>0.060</b>	<b>0.137</b>	0.008
cooking stand	0.014	0.007	0.017	<b>0.156</b>
coarse-ware jug	0.002	0.021	0.002	<b>0.047</b>

on the other hand (negative pole). It has to be noted that other objects lie on the positive side of the dimension, whose function turns out to be logically related to the objects having major contribution to the definition of that pole of the dimension: see, e.g., tray, cooking stand, coarse-ware jug. By the same token, fine-ware local jug (dinner pouring vessel) lies on the negative side of the dimension; the same holds true for big storing vessels.

As for the second dimension, if we take into account the function of the categories having major contribution to the definition of that dimension, the interpretation seems to be less clear-cut. While different types of objects are actually responsible of the definition of the dimension, their functions are nonetheless related to storing and processing activities. The positive pole of the dimension

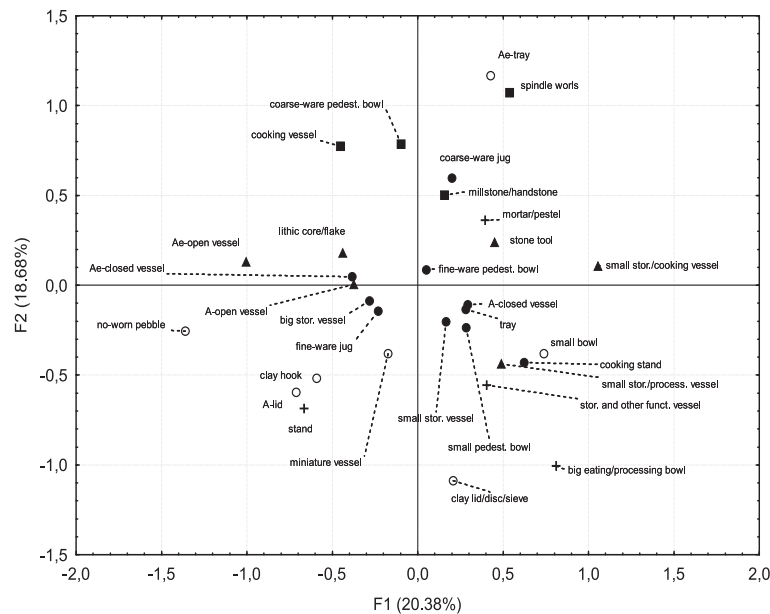


Figure 2: Symmetric scatterplot of CA performed on Table 1: row points are plotted (i.e., objects serving different purposes) on the plane defined by the first and second dimension. Solid circles: active points; hollow circles: supplementary points; triangles: points having major contribution to the definition of the first dimension only; crosses: major contributors to the first and second dimension; squares: major contributors to the second dimension only. See also Table 3

turns out to be particularly interesting, since it is defined by processing tools (mortar/pestle, millstone/handstone), cooking vessels, coarse-ware bowls, and spinning tools (spindle whorl).

After having interpreted the “meaning” of the first two dimensions, it is possible to interpret the spread and the relative distances of column points (i.e., huts). Figure 3 displays the symmetric plot showing the column points. In that figure, in order to provide right into the column map a simplified version of the information provided by the preceding Figure 2, the dimensions have been labelled according to the objects contributing to their definition (see, e.g., Bendixen, 1995). Table 4 reports the quality of the display of column points (i.e., the percentage of their inertia explained by the dimensions) and the correlation of the column points to the dimensions (i.e., square root of points’ squared cosines: Greenacre, 2007, p. 86).

Three broad groups can be isolated. The first lies in the lower right sector of the plane and is made up of huts associated with functions like working, processing, small storing, cooking, and storing/cooking. It has to be noted that some huts are not well represented on this plane; in particular, the majority of the

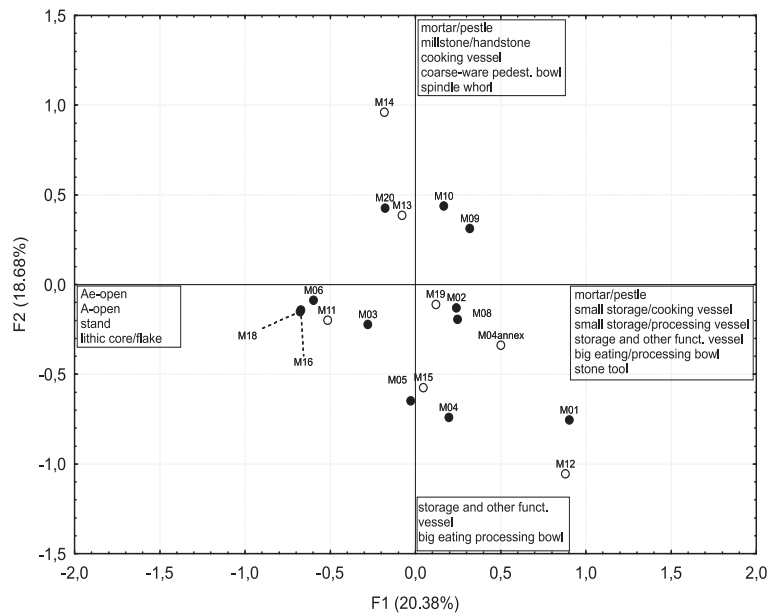


Figure 3: Symmetric scatterplot of CA performed on Table 1: column points (i.e., huts) are plotted on the plane defined by the first and second dimension; axes are labelled according to the row categories (i.e., objects) having major contribution to the definition of the dimensions

inertia of M02 and M08 is captured by the first and fourth dimension, and those huts are strongly correlated to the latter dimension. Nevertheless, it can be noted that the picture remains the same if we take into account the plane defined by the first and fourth dimensions (Figure 4). Those huts are related to functions like cooking and to other ones requiring coarse-ware vessels.

A second group can be isolated in the upper portion of the plane, and is made up of huts associated with functions like processing, cooking, and spinning. Hut M10 and M20 are well represented by the first and third dimension and have a high correlation to the latter. An inspection of the plane defined by those dimensions (Figure 5) shows that hut M10 is related to functions like storing/cooking, spinning and (in this only instance) to the use of non-local eating/drinking vessels. M20, on the other hand, is correlated to cooking and processing.

Referring back to Figure 3, the last group lies on the left sector of the plane and is made up of huts mainly associated to fine-ware imported vessels used for drinking/eating. Besides, as previously stressed, on this side of the plane objects also lie used for pouring liquids (i.e., fine-ware local jugs).

### 5.3 Pattern Interpretation and Grounds for Further Inquiries

The preceding analysis has shown how CA turns out to be a valuable tool in

Table 4: Quality of the display of column points (i.e., huts) on the planes defined by pairs of dimensions. Correlation of column points to the dimensions is also shown

Column Points	Quality of display			Correlaton to Dimensions			
	F1+F2	F1+F3	F1+F4	F1	F2	F3	F4
M01	0.636	0.390	0.375	0.61	0.51	0.12	0.02
M02	0.177	0.285	0.570	0.37	0.20	0.38	0.66
M03	0.368	0.232	0.246	0.47	0.38	0.08	0.15
M04	0.456	0.036	0.376	0.17	0.65	0.08	0.59
M05	0.410	0.018	0.062	0.03	0.64	0.13	0.25
M06	0.292	0.293	0.289	0.53	0.08	0.08	0.06
M08	0.119	0.108	0.469	0.27	0.21	0.18	0.63
M09	0.431	0.250	0.551	0.47	0.46	0.17	0.57
M10	0.327	0.653	0.058	0.20	0.53	0.78	0.13
M16	0.588	0.580	0.567	0.75	0.16	0.13	0.06
M18	0.523	0.540	0.515	0.71	0.16	0.20	0.13
M20	0.267	0.534	0.040	0.20	0.48	0.70	0.01
M04annex	0.144	0.122	0.201	0.31	0.21	0.15	0.32
M11	0.181	0.202	0.159	0.40	0.15	0.21	0.04
M12	0.168	0.069	0.160	0.26	0.31	0.01	0.30
M13	0.066	0.243	0.007	0.05	0.25	0.49	0.06
M14	0.060	0.010	0.016	0.05	0.24	0.09	0.12
M15	0.190	0.002	0.196	0.04	0.43	0.02	0.44
M19	0.019	0.040	0.014	0.10	0.09	0.17	0.07

pinpointing different activity areas at the on-site level. The pattern of distribution of functional objects across the village's structures provides grounds to make further speculations and to infer about the way in which life was organized within the past community. Obviously, the patterns highlighted by CA are to be understood and explained by the archaeologist in a way that makes sense in a "human" and social perspective. Nevertheless, it is apparent the relevance of CA as tool hallowing latent data structure to emerge.

In our case study, CA succeeded in isolating two types of huts and related artefact inventories. The first type is made up of cabins whose function turns out to be related to activities that can be labelled as utilitarian. These hosted activities related to working, storing (in small quantities), processing, cooking, and spinning. The second type of hut, which can be identified with the third aforementioned group, is made up of few cabins having relation to functions like food consumption (eating/drinking, pouring) and storing of huge quantities of goods. Remarkably, food consumption turns out to be performed by means of both local and non-local fine-ware ceramics. I would lean to label these huts as dwelling structures.

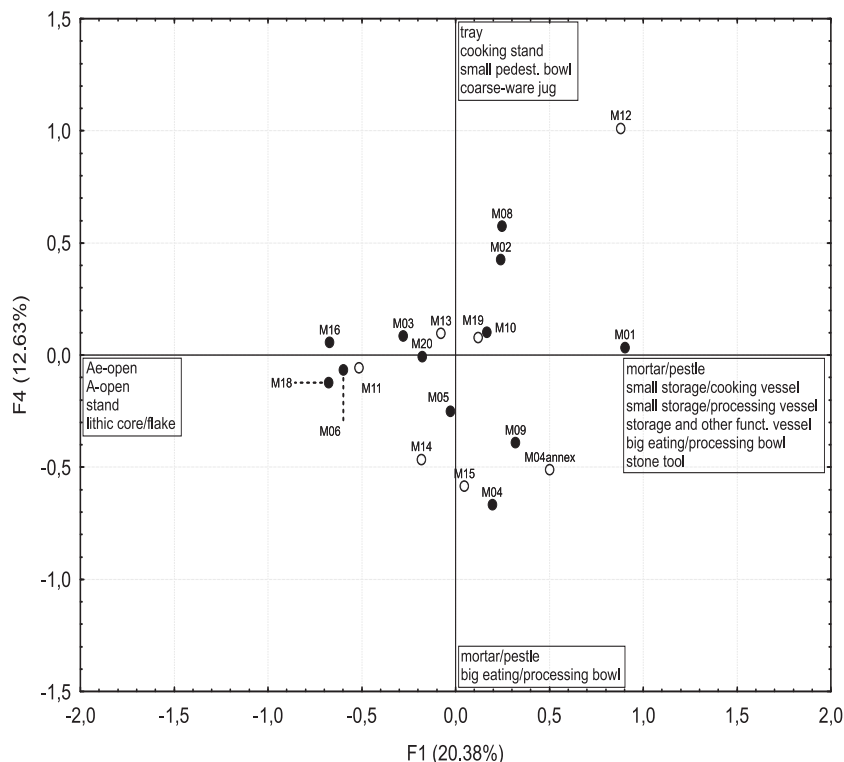


Figure 4: As Figure 3, showing the first and fourth dimension

Referring back to Figure 1D, if we look at the settlement plan where huts are given colours according to the functions stemming from the above analysis, it is apparent that the distribution of the structure across the site begins to make some sense. It could be envisaged an organization made up of few dwelling huts, each possibly coupled by at least one utilitarian huts. See, for instance, huts M16-20, M06-01, M03-02-08, M11-09, M18-10. Incidentally, it must be noted that a similar organization is not unknown in African communities object of ethno-archaeological investigations (David, 1971; Hodder, 1982, pp. 130-136; full discussion in Alberti, 2012, pp. 227-228).

Finally, it will be touched upon how the results of CA can provide grounds for further speculations. In fact, analyst could be interested in assessing: (1) to what extent a different huts' function is related to a different floor area; (2) once dwelling huts have been identified, how many people are likely to have inhabited the surviving part of the settlement.

If the huts' main room (i.e., unroofed annexes excluded) floor area is taken into account (Table 5), it is apparent that there is a significant tendency for dwelling huts to be greater than utilitarian cabins ( $U = 5$ ,  $Z = -2.499$ ,  $p = 0.014$ ) (Figure 6A).

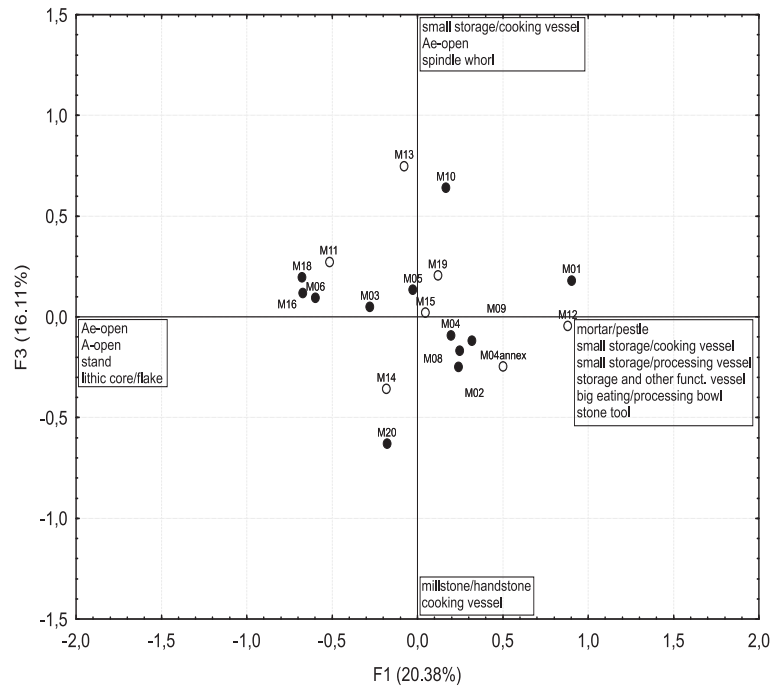


Figure 5: As Figure 3, showing the first and third dimension

Table 5: Floor area of the P. Milazzese huts' main room. Function (as hypothesised on the basis of the interpretation of CA result) and estimated number of potential residents are also shown

Huts	Floor area (sq. m.)	Function (as hypothesized after CA)	Number of potential residents (assuming 1 person every 10 sq.m.)
M 03	16.97	dwelling	1.7
M 06	18.52	dwelling	1.9
M 11	15.58	dwelling	1.6
M 16	21.13	dwelling	2.1
M 18	19.91	dwelling	2.0
M 01	9.96	utilitarian	1.0
M 02	13.57	utilitarian	1.4
M 04	19.53	utilitarian	2.0
M 05	16.35	utilitarian	1.6
M 08	10.7	utilitarian	1.1
M 09	11.92	utilitarian	1.2
M 12	10.97	utilitarian	1.1
M 13	16.65	utilitarian	1.7
M 15	10.38	utilitarian	1.0
M 20	14.11	utilitarian	1.4

An even more clear-cut picture is arrived to if we consider that the utilitarian hut having the largest floor area (M04) could have originally been a dwelling cabin that underwent a change of functional destination, as the ones documented in the ethno-archaeological literature (devolutionary reuse sensu Horne, 1994, p. 180). If M04 is dropped from the analysis, dwelling huts turn out to be very significantly larger than utilitarian ones ( $U = 2, Z = -2.733, p = 0.006$ ) (Figure 6B). Incidentally, this evidence is consistent with what is cross-culturally documented in other archaeological sites (see Alberti, 2012, p. 56, pp. 220-221 for further details).

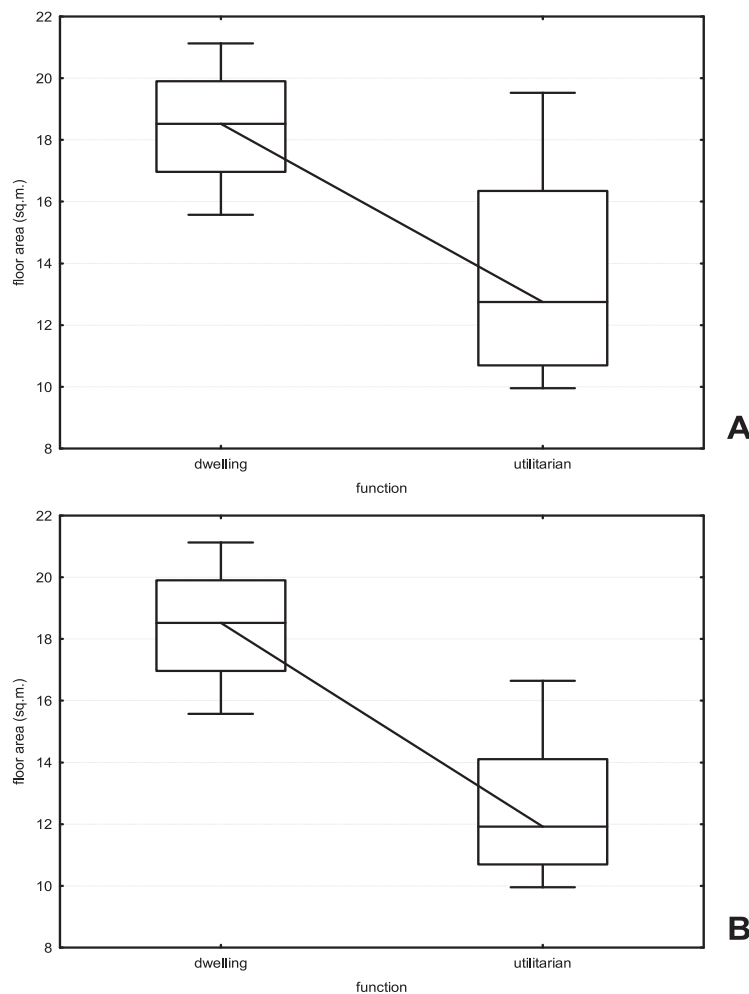


Figure 6: Boxplots of the huts' floor area (sq. m.) grouped by huts' function. A) All huts; B) hut M04 excluded. In both cases, the difference in floor area between the two groups is significant (A:  $U = 5, Z = -2.449, p = 0.014$ ; B:  $U = 2, Z = -2.733, p = 0.006$ )

As for the above second point, the identification of dwelling huts can provide grounds for estimating the number of inhabitants. Acknowledging the fact that this line of inquiry is a much-debated one in archaeology (overview in Porcic, 2012), we could consider as working hypothesis the ratio of 10 sq. m. per person as originally proposed by Naroll (1962), subsequently refined by LeBlanc (1971) who suggested to apply Naroll's finding only to the total roofed dwelling area, and recently used in studies on Neolithic communities of the Near-East (Kuijt, 2000, p. 85). It is apparent that the dwelling huts at P. Milazzese cannot accommodate more than two individuals (rounding the figures in Table 5 to the nearest whole number), that is a nuclear family. Remarkably, huts identified as utilitarian do not have floor area sufficient to accommodate that type of family, and this could be considered as evidence further supporting their non-residential destination<sup>1</sup>. On these grounds, there is reason to believe that the surviving sectors of the village are likely to have accommodated five nuclear families or, in other words, ten adult individuals.

## 6. Conclusions

By means of the discussion of a case study, this article has attempted to show the potential of CA in providing aid to the activity areas research in archaeology. The possibility to display rows and columns of contingency tables allowed to explore patterns of relationship between prehistoric huts and objects found in them, so providing the basis to pinpoint spots that were likely to be designated for different kind of activities. It has also been shown how the achievements of the analysis can contribute to shed light on the spatial organization of the settlement. Further, it has been discussed how the very possibility to isolate structures with different functions hallows to open new avenues of inquiry into other aspects of the archaeological documentation. On the basis of CA results, it has been possible to test for the existence of a difference in floor area between huts used for different functions (dwelling vs. utilitarian), and to arrive to an estimate of the number of individuals that are likely to have inhabited the village. In concluding, it should be apparent how important CA turns out to be for the interpretation of archaeological settlement data, and how the technique can prove useful in putting forward new hypotheses liable to be further explored by the archaeologists.

## Acknowledgements

This article is based on part of my doctoral dissertation titled *Social Organization and Community Practices*. Analysis for a reconstruction of the social

---

<sup>1</sup>Note that the only exceptions (huts M05 and M13) may be due to an overestimate of the original floor area because of the highly eroded perimeter wall.



structure of the Aeolian communities during the Middle Bronze Age (University of Udine, Italy, 2012). I would like to express my sincere thanks to my supervisor Prof. Elisabetta Borgna for the guidance, comments, and criticisms provided during the writing of the dissertation. Thanks are also due to Prof. Giulia Recchia (University of Foggia) and Prof. Riccardo Guglielmino (University of Salento), members of the PhD committee, for their insightful and positive comments on my work, and to Dr Maria Clara Martinelli for the kind help provided in the early stages of the study. I would like to express my gratitude to Prof. Wen-Jang Huang, Editor of the Journal of Data Science, for the patience shown during the submission and reviewing process, and the anonymous referee for the constructive and insightful comments provided, allowing to greatly improve the final version of the manuscript. I am the sole responsible of any possible errors or misunderstandings.

## References

- Adams, W. Y. and Adams, E. W. (1991). *Archaeological Typology and Practical Reality: A Dialectical Approach to Artifact Classification and Sorting*. Cambridge University Press, Cambridge.
- Alberti, G. (2011). Radiocarbon evidence from the Middle Bronze Age settlement at Portella (Aeolian Islands, Italy): chronological and archaeological implications. *Radiocarbon* **53**, 1-12.
- Alberti, G. (2012). *Social Organization and Community Practices. Analysis for a Reconstruction of the Social Structure of the Aeolian Communities During the Middle Bronze Age*. Unpublished Ph.D. Dissertation, Department of History and Preservation of the Cultural Heritage, University of Udine, Udine.
- Alberti, G. (2013). A Bayesian  $^{14}\text{C}$  chronology of early and middle bronze age in Sicily. Toward an independent absolute dating. *Journal of Archaeological Science* **40**, 2502-2514.
- Baxter, M. J. (1994). *Exploratory Multivariate Analysis in Archaeology*. Edinburgh University Press, Edinburgh.
- Bellanger, L., Tomassone, R. and Husi, P. (2008). A statistical approach for dating archaeological contexts. *Journal of Data Science* **6**, 135-154.
- Bendixen, M. (1995). Compositional perceptual mapping using chi-squared tree analysis and correspondence analysis. *Journal of Marketing Management* **11**, 571-581.

- Benzécri, J. P. (1992). *Correspondence Analysis Handbook*. Statistics, Textbooks and Monograph-V. 125, Marcel Dekker, New York.
- Bernabò Brea, L. and Cavalier, M. (1968). *Meligunìs Lipàra Vol. III. Stazioni preistoriche delle isole Panarea, Salina e Stromboli*. Flaccovio, Palermo.
- Blanco Abellàn, M. (2007). Textbooks on differential calculus in eighteenth century Europe: a comparative stylistic analysis. *Journal of Data Science* **5**, 597-612.
- Blankholm, H. P. (1991). *Intrasite Spatial Analysis in Theory and Practice*. Aarhus University Press, Aarhus.
- Blasius, J. and Greenacre, M. (1998). *Visualization of Categorical Data*. Academic Press, New York.
- Bølviken, E., Helskog, E, Helskog, K., Holm-Olsen, I. M., Solheim, L. and Bertelsen, R. (1982). Correspondence analysis: an alternative to principal components. *World Archaeology* **14**, 41-60.
- Clausen, S. E. (1998). *Applied Correspondence Analysis. An Introduction*. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-121. Sage, Thousand Oaks, California.
- Cool, H. E. M. and Baxter, M. J. (2002). Exploring Romano-British finds assemblages. *Oxford Journal of Archaeology* **21**, 365-380.
- Cutting, M. (2006). More than one way to study a building: approaches to prehistoric household and settlement space. *Oxford Journal of Archaeology* **25**, 225-246.
- David, N. (1971). The Fulani compound and the archaeologist. *World Archaeology* **3**, 111-131.
- Djindjian, F. (1985). Seriation and toposeriation by correspondence analysis. In *P.A.C.T. 11, To Pattern the Past* (Edited by A. Voorrips and S. H. Loving), 119-135. Council of Europe.
- Drennan, R. D. (2009). *Statistics for Archaeologists: A Commonsense Approach*, 2nd edition. Springer, New York.
- Greenacre, M. (2007). *Correspondence Analysis in Practice*, 2nd edition. Chapman & Hall/CRC, New York.

- Hodder, I. (1982). *Symbols in Action: Ethnoarchaeological Studies of Material Culture*. New Studies in Archaeology, Cambridge University Press, Cambridge.
- Holloway, R. R. and Lukesh, S. (1995). *Ustica I. Excavations of 1990 and 1991*. Providence/Louvain-La-Neuve.
- Horne, L. (1994). *Village Spaces: Settlement and Society in Northeastern Iran*. Smithsonian Institution Press, Washington, District of Columbia.
- Kent, S. (1987). Understanding the use of space: an ethnoarchaeological approach. In *Method and Theory for Activity Area Research: An Ethnoarchaeological Approach* (Edited by S. Kent), 1-60. Columbia University Press, New York.
- Kjeld Jensen, C. and Høilund Nielsen, K. (1997). Burial data and correspondence analysis. In *Burials & Society: the Chronological and Social Analysis of Archaeological Burial Data* (Edited by C. Kjeld Jensen and K. Høilund Nielsen), 29-61. Aarhus University Press, Aarhus.
- Kuijt, I. (2000). People and space in early agricultural villages: exploring daily lives, community size, and architecture in the late Pre-Pottery Neolithic. *Journal of Anthropological Archaeology* **19**, 75-102.
- LeBlanc, S. (1971). An addition to Naroll's suggested floor area and settlement population relationship. *American Antiquity* **36**, 210-211.
- Lowell, J. C. (1991). *Prehistoric Households at Turkey Creek Pueblo, Arizona*. University of Arizona Press, Tucson.
- Madsen, T. (1989). Seriation and multivariate statistics. In *Computer Applications and Quantitative Methods in Archaeology* (Edited by S. Rahtz and J. Richards), 205-214. BAR International Series 548, Oxford.
- Mameli, L., Barcelò, J. A. and Estevez, J. (2002). The statistics of archaeological deformation process. An archaeozoological experiment. In *Archaeology at the Interface* (Edited by G. Burenhult and J. Arvidssen), 1-17. Archaeopress, Oxford.
- Naroll, R. (1962). Floor area and settlement population. *American Antiquity* **27**, 587-589.
- Panagiotakos, D. B. and Pitsavos, C. (2004). Interpretation of epidemiological data using multiple correspondence analysis and log-linear models. *Journal of Data Science* **2**, 75-86.

- Pavúk, P. (2010). Pottery processing at Troy. Typology, stratigraphy and correspondence analysis: how do they work together? In *Analysing Pottery: Processing, Classification, Publication* (Edited by B. Horejs, R. Jung and P. Pavúk), 73-98. Comenius University in Bratislava, Bratislava.
- Peeples, M. A. and Schachner, G. (2012). Refining correspondence analysis-based ceramic seriation of regional data sets. *Journal of Archaeological Science* **39**, 2818-2827.
- Pitts, M. (2005). Pots and pits: drinking and deposition in Late Iron Age south-east Britain. *Oxford Journal of Archaeology* **24**, 143-161.
- Porcic, M. (2012). Effects of residential mobility on the ratio of average house floor area to average household size: implications for demographic reconstruction in archaeology. *Cross-Cultural Research* **46**, 72-86.
- Potter, J. M. (2000). Pots, parties, and politics: communal feasting in the American Southwest. *American Antiquity* **65**, 471-492.
- Reynolds, H. T. (1977). *Analysis of Nominal Data*. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-007. Sage, Thousand Oaks, California.
- Rowntree, D. (2000). *Statistics Without Tears: An Introduction for Non-Mathematicians*. Penguin Book, London.
- Schiffer, M. B. (1989). Formation processes of Broken K Pueblo: some hypotheses. In *Quantifying Diversity in Archaeology* (Edited by R. D. Leonard and G. T. Jones), 37-58. Cambridge University Press, Cambridge.
- Shennan, S. (1997). *Quantifying Archaeology*, 2nd edition. Edinburg University Press, Edinburgh.
- Smith, K. Y. and Neiman, F. D. (2007). Frequency seriation, correspondence analysis, and Woodland period ceramic assemblage variation in the Deep South. *Southeastern Archaeology* **26**, 47-72.
- Spaulding, A. C. (1953). Statistical techniques for the discovery of artifact types. *American Antiquity* **18**, 305-313.
- Steadman, S. R. (1996). Recent research in the archaeology of architecture: beyond the foundations. *Journal of Anthropological Research* **4**, 51-93.
- Taylor, R. (1990). Interpretation of the correlation coefficient: a basic review. *Journal of Diagnostic Medical Sonography* **1**, 35-39.

- Wallin, P. (2010). In search of rituals and group dynamics: correspondence analyses of Neolithic grave fields on the Island of Gotland in the Baltic Sea. *Documenta Praehistorica* **37**, 65-75.
- Yelland, P. M. (2010). An introduction to correspondence analysis. *The Mathematical Journal* **12**, 1-23.
- Verhoeven, M. (1999). *An Archaeological Ethnography of a Neolithic Community. Space, Place and Social Relations in the Burnt Village at Tell Sabi Abyad, Syria*. Nederlands Historisch-Archaeologisch Instituut Te Istanbul.
- VanPool, T. L. and Leonard, R. D. (2011). *Quantitative Analysis in Archaeology*. Wiley-Blackwell, New York.

Received August 10, 2012; accepted February 7, 2013.

Gianmarco Alberti  
PhD-University of Udine  
gianmarcoalberti@tin.it