Coins, Copies and Kernels - a Note on the Potential of Kernel Density Estimates

Kris Lockyear

CAA 97

Abstract

One of the more remarkable aspects of the distribution of Roman Republican coinage is the vast quantities of these coins recovered from the territory of Romania, roughly ancient Dacia. Yet more remarkable is the evidence for the contemporary copying of these coins in a manner which makes the identification of them as copies extremely difficult. Obviously, some estimation of the date of these copies, and the proportion of the Dacian assemblage which are copies, is essential in any attempt to interpret their significance in social and economic terms. In order to provide some answers to these questions, an archaeometallurgical project was organised by the author. The project sampled some 200 coins from Romania, and UK museums, which were then analysed using atomic absorption spectrometry. The statistical analysis of this data, after some initial success, proved a difficult task. This paper reviews the analyses, problems and solutions, with particular emphasis on the use of kernel density estimates in the examination and interpretation of bivariate scattergrams and maps from principal components analysis.

1 Introduction

The analysis and graphical representation of large complex data sets is a problem that has been addressed in many ways with varying degrees of success. Data reduction methods, such as Principal Components Analysis (PCA) or Correspondence Analysis (CA) are often very successful but can still suffer from crowded plots. Use of colour helps to discern structure in the data, such as groups, (Scollar *et al* 1993) but is not a perfect solution. Plotting boundaries on scatter plots or maps, perhaps derived from the results of a PCA or CA, has also been used (e.g., Goldberg and Iglewicz 1992), along with two-dimensional variations on the box-plot (Becketti and Gould 1987). Many of these methods suffer, however, from a prior assumption that the underlying distribution is regular, e.g., elliptical. In many cases, this assumption is false. An alternative is the use of Kernel Density Estimates (KDEs) which can be used to plot two dimensional 'contour plots' on bivariate maps (e.g., Bowman and Foster 1993). The application of KDEs to archaeological problems was first suggested by Baxter and Beardah (1995) who have also developed routines in matlab to perform these analyses (Beardah and Baxter 1996b), and published a number of papers on their application in archaeology (Baxter and Beardah 1996; Beardah and Baxter 1996a; Baxter *et al* 1997; see also Baxter, this volume and Beardah, this volume).

The aim of this short paper is to present an example of the use of these routines in the analysis of a complex data set, and to suggest some desiderata for the future. A fuller publication of all aspects of the statistical methodology employed and the lessons learnt will be published elsewhere. The final report on the project is to be submitted to the Romanian journal Dacia. This paper will not consider the statistical theory behind the method for which the reader is referred to the excellent book by Wand and Jones (1995).

2 The problem

One of the many remarkable aspects of the numismatic history of the late Iron Age in Romania is the evidence for the copying of Roman Republican coins by the native population. Evidence for this, in the form of coin dies, cast coins and die links, indicate the presence of these copies, but the scale of the copying has been disputed. This is because the copies are so exact that normal methods of identification cannot supply an answer (see Lockyear 1996b for a full discussion of the problem).

Obviously, the significance of these copies in the development of late Iron Age society in the region is largely dependent on what proportion of the coin assemblage are genuine Roman coins, and what proportion are locally made copies. The main influx of Roman *denarii* into the region was between c. 75 and c. 65 BC. A logical context for these copies would therefore be immediately after that date (Lockyear 1996a).

In order to estimate the proportion of copies in the total assemblage, a programme of archaeometallurgical analysis was instigated by the author in collaboration with Mathew Ponting, Clive Orton, and Gheorghe Poenaru-Bordea. In May 1992, 178 samples were obtained from *denarii*, tetradrachms of Thasos, and from two silver bars found with the Stăncuţa hoard (STN[1]; Preda 1958). Amongst the *denarii* sampled were known imitations, cast and struck copies[2]. Details of the hoards and samples are given in Table 1. Subsequently, comparative material from the Ashmolean and the British Museum was sampled. The samples were analysed by Matthew Ponting using atomic absorption spectrometry and the data passed to myself for statistical analysis. A preliminary batch of 30 coins was analysed in 1992 (Lockyear and Ponting 1993), and the remaining coins in 1994-5.

The first batch of samples were analysed using a single solution method which proved to be problematic; the second batch, therefore, was analysed using a two solution method. Three samples from the first batch were re-analysed in the second batch.

Hoard	No.	Sample	Reference	Reason
Zătreni	41	6	ZAT; Chițescu 1981, no. 215	early hoard in Muntenia
Poiana	152	20	IPO; Chițescu 1981, no. 148	hoard from major settlement in Moldavia
imitations	–	6	Chițescu 1981, nos. 11, 28, 84, 67, 165, 239	unprovenanced, for comparison to hoard material
Popești	?	3	in preparation	3 tetradrachms of Thasos, by request of Poenaru-Bordea
Breaza	122†	19	BRZ; Poenaru Bordea & Știrbu 1971; Chițescu 1981, no. 29	contains cast copies
Stăncuța	34	9	STN; Preda 1958; Chițescu 1981, no. 188	mixed hoard of tetradrachms, *denarii* and silver bars
Voinești‡	94	3	VOI; Știrbu 1978, p. 90, no. 4	by request of C. Știrbu
Poroschia	552	66	PRS; Chițescu 1980; Chițescu 1981, no. 154	contained possible copies
Șeica Mică	348	44	SEI; Floca 1956; Chițescu 1981, no. 193	hoard from Transylvania, used by Crawford in RRC

Table 1. Romanian hoards sampled May 1992. †București lot. ‡Not published in detail and therefore contents not listed in the CHRR database.

3 Analysis

The data analysis had a number of stages:

1. Data cleaning;
2. Univariate analysis using dot-plots and summary statistics;
3. Comparison of elements to the date of minting;
4. Bivariate analysis using scattergrams and KDE 'contour' plots;
5. Multivariate analysis using PCA, the results of which were plotted on a map along with contours derived from KDE;
6. Estimation of the number of copies per hoard using the results from 5.

This paper will concentrate on stages 4 and 5, but will first summarise briefly the other stages.

3.1 Stages 1-3

The data initially required some checking and cleaning to remove erroneous data points and measurements, to identify analyses undertaken on very small samples, etc. To do this the data were converted from a series of excel spreadsheets to a relational database structure. To account for variable sample size a dual method of estimating missing data values was used (Lockyear 1996b, 410-15).

One of the major problems with the data set is that the silver alloy used is extremely pure. This meant that although some 13 elements were looked for, very few were detectable in the majority of cases. The univariate analyses used both summary statistics (Lockyear 1996b, Table 14.9), and dot-plots (Lockyear 1996b, Figs. 14.9-14.17) to examine each element individually, and specific coins which appeared unusual on this basis were noted. These analyses identified one of the coins from the Ashmolean Museum as being a copy! This was a timely reminder that simply because a coin comes from a UK museum does not mean it is necessarily genuine.

Of the elements looked for, only five had sufficient measurements above the detection limit for further analysis. Of these, silver formed over 93% of 75% of the coins. In order to avoid problems of closure (Aitchison 1986)--that is the fact that the all the percentages sum, obviously, to 100--this element was dropped from all subsequent analyses[3]. This left four elements: copper (Cu), lead (Pb), gold (Au) and bismuth (Bi). Each of these elements were plotted against the date of the coins (Lockyear 1996b, 421-24) to test for possible temporal trends in the data. There was the slightest hint that copper levels may have increased from 157 to 50 BC, but otherwise no temporal patterning was detected.

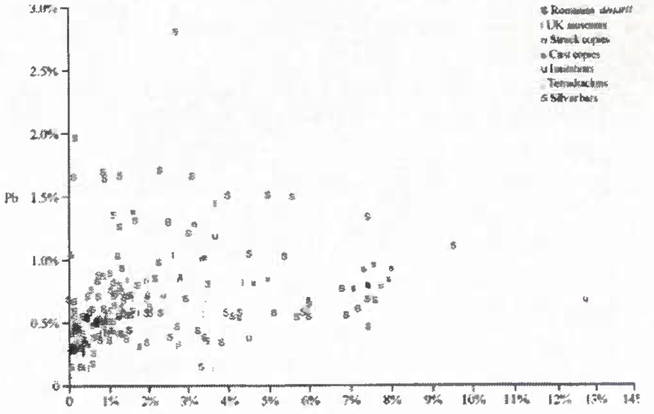

Figure 1. Scattergram of Cu v. Pb for all samples (except 191 which was omitted due to poor data)

3.2 Stage 4 - bivariate analysis

Two bivariate scattergrams were constructed; one of the two major elements (Cu v. Pb) and one of the two minor elements (Au v. Bi). The use of multiple symbols allowed some grouping to be seen in the plots, e.g., the cast coins from Breaza (BRZ) fall into a small group, but the pattern is far from clear (Fig. 1). It would seem that the copies generally have higher levels of copper and/or lead, and that most UK museum coins have low levels of these elements, but the separation is not clear cut.

In order to make the division between known copies, UK museum coins (assumed to be mainly genuine) and the remainder of the *denarii* clearer, a number of KDE contour plots were created -- see Figure 2 for an example. These plots were created with the kdedemo2 set of macros for matlab which will be discussed in section 4.1. The package offers four different kernel functions and a variety of methods for estimating the value of the bandwidth h; this procedure is analogous to selecting the bin-width when constructing a histogram (Beardah and Baxter 1996b). These different options were tried, but the 'best' results, judged solely on visual criteria, were obtained by using the recommended options of the normal kernel and using the Sheather-Jones method of selecting h (called 'solve-the-equation 2' in the kdedemo2 package).

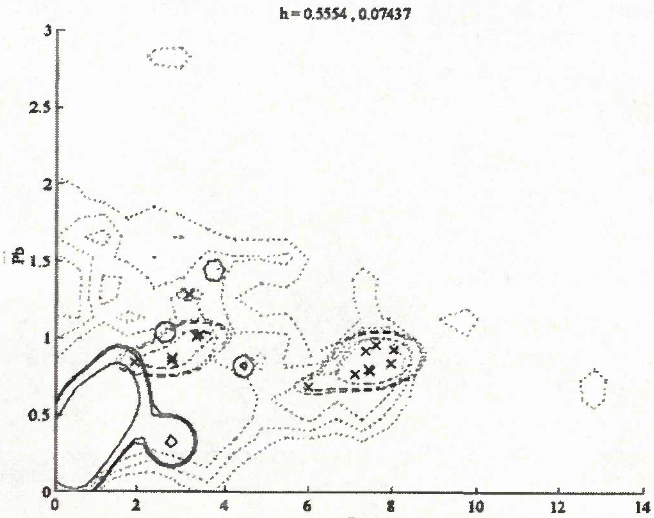

Figure 2. Kernel density estimate percentage contour plot, 85, 95 and 100% contour lines for: all samples (dotted), UK museums (solid) and cast/struck copies (dashed). Sample 191 omitted. Cu v. Pb. Crosses mark location of cast/struck copies.

Figure 2 plots the 85%, 95% and 100% contour levels for the three groups of coins. In lay terms, the 85% 'contour' line for copies contains 85% of the data points for those copies whilst maximising the density of data points within that line. This principle obviously applies to each line/group. It can be clearly seen from the figure that there is good separation between the UK museum coins (bottom left) and two groups of copies. These two groups represent the cast coins from Breaza (the right- hand group) and the struck copies from Poroschia (left-hand group). There is a slight overlap with one copy lying on the edge of the main group of UK coins, and three UK coins lying away from that main group. It is important to note that there are many other 'unknown' *denarii* that lie outside the main UK group, interspersed and surrounding the copies. Similar patterning was observed with the two minor elements.

3.3 Stage 5 - the multivariate analysis

Success with the bivariate plots suggested that a multivariate analysis might increase the separation between groups. PCA was performed on the data set using the same four elements. The analysis was performed using a correlation matrix and the first two axes 'explained' 59.1% of the variation in the data.

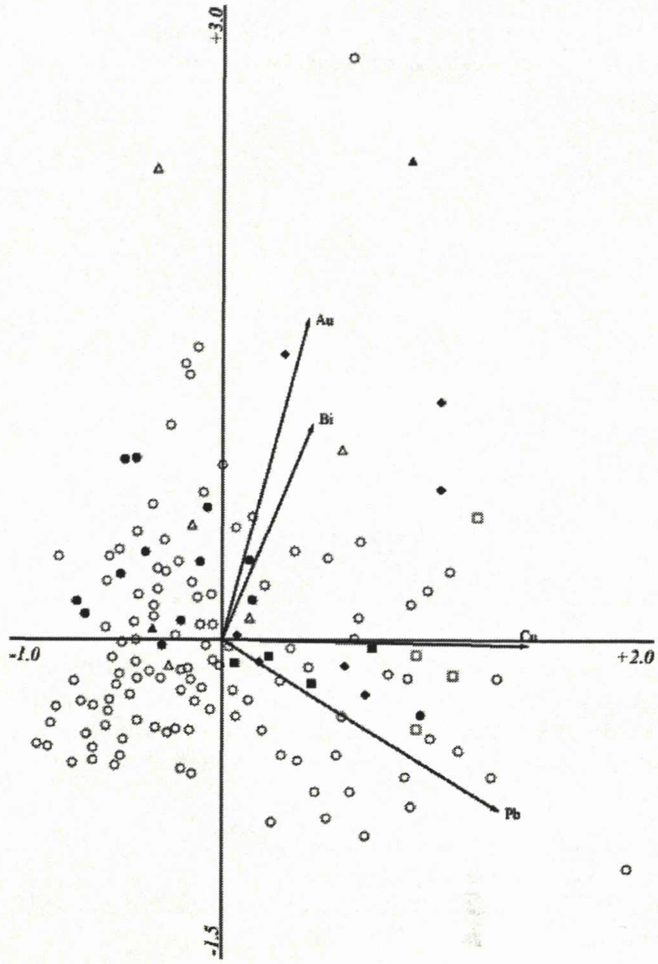

Figure 3. Biplot from PCA of full metallurgical data set omitting sample 191. 1st and 2nd axes of inertia. Open circles: *denarii* from Romania; filled circles UK museum *denarii*; open squares: cast copies from Breaza; filled squares: struck copies from Poroschia; open triangles: tetradrachms; filled triangles: silver bars; diamonds: imitations.

Figure 3 presents the biplot from this analysis. As can be seen, there is a correlation between copper and lead, and a second correlation between gold and bismuth. The first principal axis mainly represents variation in the copper levels, and to a lesser extent the lead values; the second axis represents the gold and bismuth values which appear to be moderately negatively correlated with lead. As can be seen from the plot, the majority of the UK museum samples occur in the top left quadrant of the plot, i.e., they are associated with low levels of all four elements. This of course means that they are actually associated with high levels of silver and thus the problem of closure has not been completely avoided by dropping that element.

Again, however, the patterning is not completely obvious. Some points are clear, for example the three data points that lie at the top extreme of the second axis are all from the Stăncuţa (STN) hoard, and all have high levels of gold. What makes this

even more fascinating is that these points represent three different types of object: a tetradrachm of Thasos, a Republican *denarius*, and a silver bar.

To make the pattern clearer a further set of KDE contour plots were produced using the normal kernel function and the Sheather-Jones method of selecting h of which Figure 4 is an example. In this plot only *denarii* were included in the calculations, and the separation of the UK museum coins and the copies can be clearly seen. There is still a little overlap, but it is minimal[4].

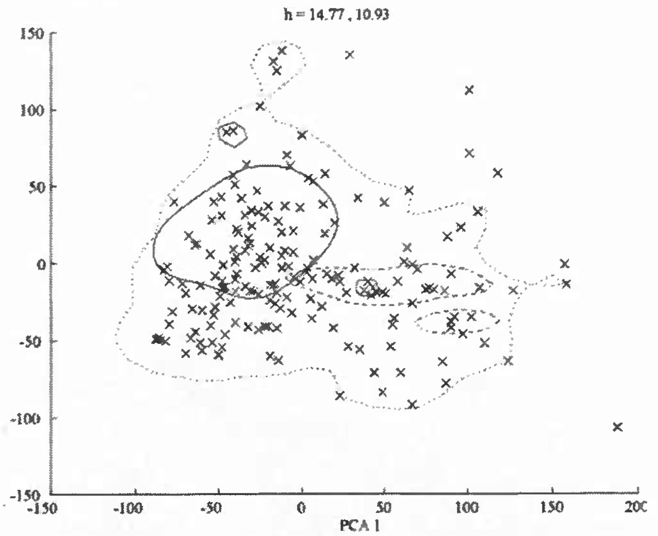

Figure 4. Kernel density estimate percentage contour plot of PCA 1 v. PCA 2. 95% contour lines marked: all *denarii* (dotted), UK museums (solid) and cast/struck copies (dashed). Individual points also marked. Samples 58 & 191 omitted.

The difficulty now was to divide the unclassified *denarii* from Romania into two classes: those which are probably copies and those which are probably genuine. To do this the *denarii* were divided into four groups. These groups were labelled 'core', 'penumbra', 'outside' and 'far-out'. The first group, the core coins, consisted of all coins which lay within the main 95% contour line for the UK museum material in Figure 4. The far-out group consisted of all coins which had a score of >0.5 on the first principal axis and >1.15 or <-0.35 on the second. These limits excluded all UK museum coins. The three coins analysed twice (see section 2) showed a 0.25-0.31 difference in their co-ordinates, mainly on the first principal axis. The coins that were left were therefore divided into the penumbra category which included all coins within 0.31 of the main 95% UK coin contour, and the outside group which were all coins that lay outside that band but within the limits set for the far-out category. A number of observations can be made about the composition of these four groups. Firstly, only a single UK museum coin lies in the far-out category, and a further coin in the out-side category. In the case of the former (coin no. 181) this is the copy identified as such by the univariate analyses discussed above. Of the cast coins from Breaza, all lie in the far-out category, but the struck coins from Poroschia mainly lie in the penumbra or outside categories. The imitations occur in most groups including the core group.

3.4 Stage 6 - estimating totals of copies

The final stage was to estimate the number of copies per hoard. To do this I decided to use the number of coins in the far-out category as the number of copies. No doubt some of the coins in this group are in fact genuine, but there is also a likelihood that some of the coins in the core category are copies as shown by the fact that some of the imitations have the same metallurgical composition as the main mass of points.

Estimates and confidence intervals were obtained using the method outlined by Shennan (1988) for those hoards where enough samples were taken. For the remainder the confidence limits were obtained from Table P of Rohlf and Sokal (1995). For the Şeica Mică (SEI) hoard 16 coins from the 44 sampled fell in the far-out category giving an estimate of 36.3% with 95% confidence limits of ± 12.4%. This seems remarkably high. Taking an ultra-conservative view and only accepting coins with very high copper levels we still get six copies, giving us 13.6% ± 8.8%. For Poroschia (PRS) 30 coins from 62 were in the far-out category giving us a figure of 48% ± 11%, or 16% ± 8% if we take the ultra-conservative line. Similar figures were obtained from the other hoards.

From this it can be seen that although the precise level of copying is still very unclear, and will become the subject of a further project, it can no longer be sustained that the level of copying in the region was minimal and insignificant (cf. Crawford 1980; Crawford 1985). This phenomenon appears to be a major part of the numismatic history of the region, and certainly reflects important political and social developments at that time.

4 Discussion

4.1 The kdedemo2 package

The kdedemo2 set of macros for matlab were written by Beardah (Beardah and Baxter 1996a; Beardah and Baxter 1996b). These macros were downloaded from Nottingham Trent University in April 1996 and were run on a UNIX workstation although they would work equally well on a Windows-based PC. They are available via Internet Archaeology.

A few comments about matlab and these macros are appropriate here. Firstly, the PostScript produced by matlab is not true encapsulated PostScript and required some manual editing by the author before the plots were usable by other packages. Secondly, the differences between line types were sometimes poorly rendered on the printed plots. As regards the kdedemo2 macros, these were easy to use and very powerful. One facility which the author would have liked was the ability to calculate/plot the contours for one group, and then list the data-points which lay outside that contour from a second group; facilities were provided to list outliers from a single group. It is possible, however, to obtain this information direct from matlab (Beardah *pers comm*). Also, a check box to force matlab to create true maps where $x = y$, essential in the plotting of the results of CA, would be useful.

The division of the coins into the four categories was undertaken manually from the plots and other data. The additional features suggested for the kdedemo2 macros would have made this task easier, and it is hoped that they might be implemented at a later date. A more experienced user of matlab would have been able to use the full potential of the package to speed-up this task.

4.2 Final comments

From the above it can be seen that KDEs are a powerful exploratory tool, particularly when used with PCA or CA of complex data sets. The method enabled the successful investigation of the problem. In terms of the current example, it would be useful if some more formal method could be developed which would integrate the metallurgical data along with other information, such as the date of the coin, the weight and the diameter, to provide a estimate of the probability for each coin being a copy. It has been challenging and I hope useful to investigate the different methods, in particular the use of KDEs, by which this very complex data set can be analysed. Nevertheless, the most challenging and interesting question remains to be answered. Why did the Dacians copy Roman Republican *denarii*?

Acknowledgements

I would like to thank my close collaborators on this project, Matthew Ponting and Clive Orton. Matthew took and analysed all the samples used in this analysis for which I am very grateful, as well as patiently answering my innumerable questions about the technique and the data. I would also like to thank John Merkel for allowing the first 30 samples to be analysed. Grants were generously awarded by the Fund for Applied Science in Archaeology, the Royal Numismatic Society and the UK Numismatic Trust, which enabled the completion of the AAS analyses. I would also like to thank my colleagues in Romania, Gh. Poenaru-Bordea, Adrian Popescu and his family, Radu Ocheşeanu, Constanţa Ştirbu, Catalin and Geta Borţun, Ecaterina Ţinţăreanu, Aniko Mitruly, Lili Ardea, Doina Comşa and Virgil Mihăilescu-Bîrliba and family. Without their help and support this project would have been impossible. I would also like to thank the British Museum, particularly Roger Bland and Andrew Burnett, and the Heberdean Coin Room, Ashmolean Museum, particularly Chris Howgego and Cathy King, for allowing us to sample coins from their collections. Mike Baxter and Christian Beardah are to be thanked firstly for bringing KDEs to my attention, but also for their help and encouragement. I have also had many long and fruitful discussions on this problem with Richard Reece, to whom I am grateful. Finally, Clive Orton, Mike Baxter, Christian Beardah and Dale Serjeantson kindly read drafts of this paper. Any errors which remain are, as always, my own.

Notes

1 All hoards mentioned in this text are accompanied by a three letter code in small capitals. This is a unique identifier used in the author's Coin Hoards of the Roman Republic CHRR database, and in other publications and allows the precise identification of individual hoards.

2 The cast copies from the Breaza (BRZ) hoard were deliberately selected and therefore have to be left out of any estimation; some coins in the Poroschia (PRS) hoard were suggested by Chiţescu to be copies, but were not deliberately selected for. In 1993, subsequent to sampling but prior to analysis, these coins were positively identified as struck copies by the author on the basis of die-linkage.

3 This method has been called naïve but in practice is a robust option and does not suffer from the problems which often occur with the alternatives (see Tangri and Wright 1993 and the reply by Baxter 1993).

4 Note that this plot has been `clipped' to exclude sample 58 which although included in the calculation of the contour lines, was so extreme it caused the plot to be very compressed.

Bibliography

Aitchison, J A, 1986 The Statistical Analysis of Compositional Data. Chapman and Hall, London

Baxter, M J, 1993 Comment on D. Tangri and R. V. S. Wright, 'Multivariate analysis of compositional data..., Archaeometry, 35 (1)(1993)', Archaeometry 35(1), 112-115

Baxter, M J and Beardah, C C 1995 Graphical presentation of results from Principal Components Analysis, In Huggett, J and Ryan, N S (eds.), Computer Applications and Quantitative Methods in Archaeology 1994, 63-67, British Archaeological Reports International Series No. 600, Oxford

Baxter, M J and Beardah, C C, 1996 Beyond the histogram: improved approaches to simple data display in archaeology using Kernel Density Estimates, Archeologia e Calcolatori, 7(1), 397-408

Baxter, M J, Beardah, C C and Wright, R V S, 1997 Some archaeological applications of Kernel Density Estimates, Journal of Archaeological Science, 24(4), 347-354

Beardah, C C and Baxter, M J, 1996a The archaeological application of Kernel Density Estimates, Internet Archaeology 1, http://intarch.ac.uk/journal/issue1/beardah index.html

Beardah, C C and Baxter, M J, 1996b MATLAB routines for Kernel Density Estimation and the graphical representation of archaeological data, In Kamermans and Fennema (1996), 179-184

Becketti, S and Gould, W, 1987 Rangefinder box plots. a note, The American Statistician, 41(2), 149

Bowman, A and Foster, P 1993 Density based exploration of bivariate data, Statistics and Computing, 3, 171-177

Chiţescu, M, 1980 Tezaurul de la Poroschia (jud. Teleorman) şi unele probleme privind monedele geto-dacice de tip roman republican, Studii şi Cercetări de Numismatică 7, 53-70

Chiţescu, M, 1981 Numismatic Aspects of the Dacian State. British Archaeological Reports, Oxford, International Series 112

Crawford, M H, 1980 Imitation of Roman Republican denarii in Dacia, Studii şi Cercetări de Numismatică 7, 51-2

Crawford, M H, 1985 Coinage and Money under the Roman Republic, Methuen, London

Floca, O, 1956 Contribuţii la cunoaşterea tezaurelor de argint dacice: tezaurul de la Sărăcsău şi Şeica Mică Editura Academiei Republicii Populare Romîne, Bucureşti

Goldberg, K M and Iglewicz B, 1992 Bivariate extensions of the boxplot, Technometrics, 34(3), 307-320

Kamermans, H and Fennema, K (eds.) 1996 Computer Applications and Quantitative Methods in Archaeology CAA95, Institute of Prehistory, University of Leiden, Leiden, Analecta Praehistorica Leidensia 28

Lockyear, K, 1996a Dmax based cluster analysis and the supply of coinage to Iron Age Dacia, In Kamermans and Fennema (1996), 165-178

Lockyear, K, 1996b Multivariate Money. A statistical analysis of Roman Republican coin hoards with special reference to material from Romania, Ph.D. thesis, Institute of Archaeology, University College London

Lockyear, K and Ponting, M J, 1993 Metallurgical Analysis of Roman Republican denarii in Romania -- Interim Report, Unpublished

Poenaru Bordea, Gh and Ştirbu, C, 1971 Tezaurul de denari romani republicani şi de la începutul principatului descoperit la Breaza, comuna Lisa (jud. Braşov), Studii şi Cercetări de Numismatică 5, 265-282

Preda, C, 1958 Asupra descoperirii monetare de la Stăncuţa (Reg. Galaţi), Studii şi Cercetări de Numismatică 2, 239-251

Rohlf, F J and Sokal, R R, 1995 Statistical Tables W. H. Freeman and Company, New York, third edition

Scollar, I, Herzog, I and Greenacre M J, 1993 Colour and graphic display aids for correspondence analysis, In J. Andresen, T. Madsen and I. Scollar (eds.), Computer Applications and Quantitative Methods in Archaeology 1992, 325-327, Aarhus University Press, Aarhus

Shennan, S J, 1988 Quantifying Archaeology, Edinburgh University Press, Edinburgh

Ştirbu, C, 1978 Noi tezaure monetare intrate în patrimoniul Muzeului de Istorie al R. S. România, Cercetări Numismatice, 1, 89-92

Tangri, D and Wright, R V S, 1993 Multivariate analysis of compositional data: applied comparisons favour standard principal components analysis over Aitchison's loglinear contrast method, Archaeometry 35(1), 103-115, (Pages 112-115, a reply by M. J. Baxter)

Wand, M P and Jones, M C, 1995 Kernel Smoothing. Monographs on Statistics and Applied Probability 60, Chapman and Hall, London

Contact details

Kris Lockyear
University College London
The Institute of Archaeology
31-34 Gordon Square
London
WC1H 0PY
email: tcfacll@ucl.ac.uk