# 18

# Statistical analysis of ceramic assemblages — a year's progress

Paul Tyers

Clive Orton

*(University College London, Institute of Archaeology, 31–34 Gordon Square, London WC1H 0PY)*

## 18.1 Introduction

This is the second interim report on the SERC-funded project of this name, and follows directly on from the first (Orton & Tyers 1989). The year has been spent in building on the theory presented then, adapting it where necessary to the peculiarities of archaeological data, and exploring the implications for methods of cataloguing pottery and defining fabric and form types. The term PIE (= pottery information equivalent) has been adopted as the basic unit of quantity of pottery after statistical transformation: a PIE is the quantity of pottery that would give as much information (in the statistical sense) about the composition of an assemblage as one whole vessel. The various aspects of the statistical analyses have been integrated into a computer package PIE-SLICE, which is being tested.

## 18.2 Theory

Several theoretical and practical problems have been overcome in the development of PIE-SLICE, mainly in the area of analysing sparse contingency tables. We start from the general theory of quasi-log-linear models (Bishop *et al* 1975, pp. 177–228), with the subscripts i, j and k usually referring to context, fabric and form respectively. The nine nested models which form our hierarchy are:

**Model 1:** no interaction between context, fabric or form, called [1][2][3].
**Model 2a:** interaction between fabric and form only, called [23][1].
**Model 2b:** interaction between context and form only, called [31][2].
**Model 2c:** interaction between context and fabric only, called [12][3].
**Model 3a:** interactions between fabric and form and between context and form only, called [23][31].
**Model 3b:** interactions between fabric and form and between context and fabric only, called [12][23].
**Model 3c:** interactions between context and form and between context and fabric only, called [31][12].
**Model 4:** all pairwise interactions, called [12][23][31].
**Model 5:** the saturated model, called [123].

The relationships between these models are illustrated in Fig. 18.1.

The fits of the various models can be examined using the likelihood-ratio statistic $G^2$ (Bishop *et al* 1975, p. 125). However, the approach cannot be used without modification because of the sparse nature of typical datasets. Three modifications were found to be necessary:

1. rules for the treatment of zero entries,
2. reduction of the size of the dataset,
3. comparisons of the fits of different models.

### 18.2.1 Random and structural zeros

Many of the entries, typically over half, are zeros. In quasi-log-linear analysis, they are treated differently according to whether they are considered to be random or structural zeros. The conventional definitions are:

> A **structural zero** is a value which is zero *a priori*, i.e. it could not possibly take a positive value (example: number of samian amphorae).
> A **random zero** is a value which is zero, but not *a priori*, i.e. it could have taken a positive value, but does not do so in our particular dataset.

Three approaches were considered:

1. treat all zeros as random, i.e. use log-linear models. This leads to high numbers of degrees of freedom, and frequently to a situation in which almost any model would fit the data.
2. treat zeros as structural if they correspond to a zero in a marginal two-way table. This approach ('conditioning on all the data') gives realistic degrees of freedom but obscures the points which may be of most interest archaeologically.
3. treat zeros as structural if they correspond to a zero in a marginal two-way table which has already been shown to relate to a significant interaction ('conditioning on the model'). This gives rise to fewer degrees of freedom than 1 but more than 2; more importantly, it seems to correspond most closely to archaeological reality.

The third method was adopted. Although the best of the three, it did not entirely solve the problem of the dilution of significant effects by large numbers of 'small' cells nor that of the occasional erratically large contribution to $G^2$ due to a very small fitted value.

### 18.2.2 Reduction of the size of datasets

This is achieved by use of the new technique SRD ('simultaneous reduction of dimensions', see Orton & Tyers, this volume). We note that because SRD works on two-way tables, the reduction of a three-way table proceeds in two stages

1. reduction of a chosen marginal table

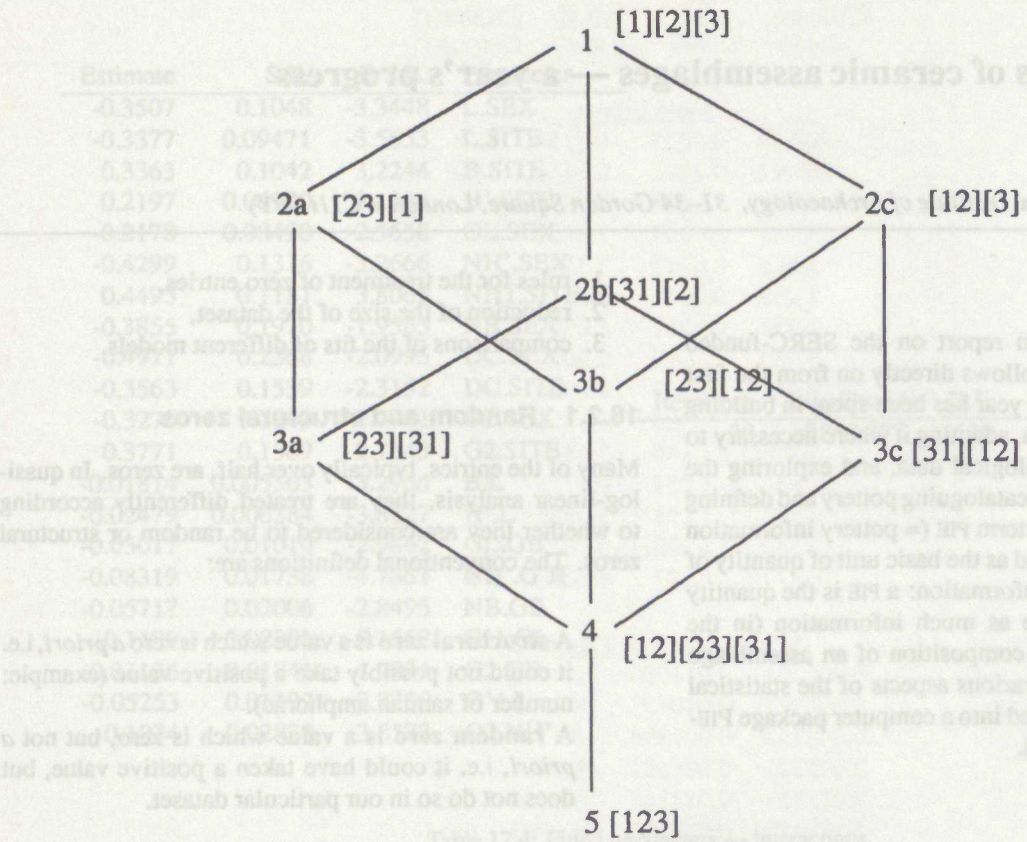


Figure 18.1: Hierarchy of models

2. reduction of the two-way table formed by treating the entries of the table created by 1 as marginal totals. This is called 'double-reduction.'

### 18.2.3 Comparative analysis of hierarchies of models

We break the hierarchy of models (Fig. 18.1) into three sub-hierarchies:

A: models 1, 2a, 3a and 3b, 4, 5;
B: models 1, 2b, 3a and 3c, 4, 5;
C: models 1, 2c, 3b and 3c, 4, 5.

We then run (quasi-)log-linear analysis on the doubly-reduced table of model IIA, (see Orton and Tyers, this volume, page 123) using sub-hierarchy A, on the doubly-reduced table of model IIB, using sub-hierarchy B, and on the doubly-reduced table of model IIC, using sub-hierarchy C.

At this point we have to take into account the need (see e.g. Craddock & Flood 1970) to ensure that the expected value in each cell is at least 1. We do this by 'flagging' all cells whose expectations fail to meet this level. Any value of one of the three variables whose 'slice' of the three-way table consists entirely of flagged cells is omitted from the analysis.

When the values of $\chi^2$ and $G^2$ are calculated, the contributions of the flagged cells remaining in the table (which we might call 'structural near-zeros') should, in principle, be ignored, as they have 'small' expected values. The question of how to count the degrees of freedom then arises.

The ideal approach would be to constrain $m_{ijk} = x_{ijk}$ for all flagged cells, in which case the fitted values would differ from those conventionally calculated, and the degrees of freedom would be reduced by the number of flagged cells (by analogy with the approach used with structural zeros). However, this approach is not followed here as it would cause great difficulties in programming.

If the differences between the observed and fitted values for the flagged cells are simply ignored, then the number of parameters fitted is the same as it would be in the conventional case, and there is no justification for reducing the degrees of freedom. Since we have reduced the overall value of $\chi^2$ (by omitting some contributions) but have kept the same degrees of freedom, this approach is statistically conservative, i.e. it will tend to miss significant patterns. There is thus a case for retaining the contributions of any remaining flagged cells. However, there is a risk of a spuriously large contribution from a cell with a very small fitted value. To avoid this risk, when a fitted value (on the current model) is $< 1$ and the corresponding observed value is $> 1$, we replace the fitted value with the value 1 for the purpose of calculating its contribution to $\chi^2$ (or $G^2$), and use all the cells. The effect is to cut back any spurious values to a more reasonable level, while making little change to 'ordinary' values. It is probably a slightly conservative procedure.

There is a hidden problem here. In the two-way case there is only one way to calculate the 'expected' values, and the criterion $x_{i.}x_{.j} > x_{..}$ (strictly speaking, $x_{i.k}x_{ij.} > x_{i..}$

in this example) can be used without ambiguity. In the three-way case, however, there are as many sets of expected values as there are models, and it is not obvious which should be used. Two possibilities are:

- model 1, giving $x_{i..}x_{.j.}x_{..k} > (x_{i..})^2$
- model 2a, giving $x_{i..}x_{.jk} > x_{...}$

It appears necessary to use the same criteria throughout any sub-hierarchy, because otherwise differences in goodness-of-fit between the models would be confused with differences between datasets. Provisionally, model 1 will be used to give the criteria, but this is an empirical decision and may need to be revised in the light of experience.

### Collapse of one or more variables into a single group

The question may arise — how do we integrate SRD if it results in the collapse of one or more variables (context, fabric and form) into a single group? The initial reaction was to reject any sub-hierarchy which led to such a collapse, but this treatment is now seen to be unnecessarily harsh. Provided only one variable collapses in a sub-hierarchy, some information can still be retrieved, as follows:

We treat the collapsed table as an ordinary two-way table, and apply standard log-linear analysis. There are no structural zeros, so a quasi-log-linear approach is not needed. The only test is of Model 1 (no interaction) against an alternative of Model 2a, 2b or 2c (i.e. the appropriate two-way interaction), depending on which variable has collapsed. If Model 1 fails, the appropriate Model 2 is automatically preferred. The test can be a simple two-way chi-squared test, or the corresponding $G^2$ test. The latter is preferred on grounds of compatibility with other tests undertaken at this stage.

This approach can also be used on datasets which are only ever two-way. In such cases, it would be preceded by a single SRD.

### Finding the 'best' model

It would be too simple to look for the 'best' overall model, since the different hierarchies may be telling us different things. For example, model 2b might be the accepted model for one grouping of contexts, while model 2c might be accepted for a completely different grouping. This would indicate form-by-context and fabric-by-context interactions cross-cutting each other, suggesting functional differences between some fabrics but chronological differences between others. To choose one model as 'best' would lose one of these interpretations.

It may be more useful to treat the outcome of the integrated program as a set of statements about interactions rather than as a set of models. Two models are 'preferred' for each hierarchy; only certain combinations are possible. Any hierarchy may give two preferences for model 1 or two for model 5 (models 1 and 5 cannot be preferred alongside any other model in the same hierarchy). The other possibilities are:

**hierarchy A :** models 2a, 3a, or 4 *and* models 2a, 3b or 4,
**hierarchy B :** models 2b, 3c, or 4 *and* models 2b, 3a or 4,
**hierarchy C :** models 2c, 3b, or 4 *and* models 2c, 3c or 4.

We call the interpretation of these outcomes the 'fruit machine' problem. For each hierarchy, we find the smallest set of interactions that are *needed* to fit the data (here called 'locally necessary interactions' or LNI's). For example, the full possibilities for hierarchy A are:

| models | | | LNI |
|---|---|---|---|
| 2a | and | 2a | [23] |
| 2a | and | 3a | [23] |
| 2a | and | 3b | [23] |
| 2a | and | 4 | [23] |
| 3a | and | 4 | [23][31] |
| 4 | and | 3b | [23][12] |
| 3a | and | 3b | [23][31] or [23][12]† |
| 4 | and | 4 | [23][31][12] |

†choose the set with the highest probability level ('best fit')

So A, B and C give LNI's of:

- none or [23] or [23][31] or [23][12] or [23][31][12] or [123]
- none or [31] or [31][12] or [31][23] or [31][12][23] or [123]
- none or [12] or [12][23] or [12][31] or [12][23][31] or [123]

respectively.

We note a slight asymmetry in that we *expect* [23] but are not especially interested in it; but we are looking particularly for [31] and/or [12].

We have thus three sets of locally necessary interactions, each arising from a *different* reduction of the original data. Each two-way interaction may be locally necessary in up to three hierarchies: we say they are either 'globally necessary' (in all three), 'sometimes necessary' (in any two), 'locally necessary' (in only one), or 'not necessary' (in none).

There is thus no simple interpretative formula that can be used. To obtain the greatest amount of interpretative information from the data, we need:

1. the partition of $G^2$ table for each hierarchy,
2. for each interaction, how 'necessary' it is,
3. for each occurrence of it,
   (a) the membership lists of the variables,
   (b) the reduced data table,
   (c) the corresponding table of expected values,
   (d) the corresponding table of deviations (i.e. observed — expected),
   (e) the contributions to $G^2$.

There is an outstanding problem — a locally necessary interaction may be really necessary, or just a product of the statistical procedures. We need criteria to distinguish between the two.

### Membership lists

The previous section raises the question — if two hierarchies partition a variable (e.g. context) in different ways, how similar are the two partitions? In answering this question, we must remember that

- the numbers of cells in the partitions may well differ
- different values may be deleted in different partitions.

We define a *link* as a pair of values in the same cell of a partition, including the link between a value and itself. We can then define the similarity coefficient S for a pair of partitions by

> S = (number of links in *both* partitions)/(number of links in *either* partition)

This coefficient appears to behave how one would intuitively hope it would. It can be used as an aid to interpretation.

We note that pairs of values may be brought together in all 3, or 2, or 1, or none of the reductions, thus measuring in some sense the universality of their similarities (e.g. do two contexts have similar fabric and/or form and/or fabric-by-form profiles?). Groupings which occur in *all* reductions are called *irreducible*. They are of particular value for interpretation.

### 18.2.4 Summary

The procedure of transformation to PIEs, reduction by SRD followed by (quasi-)log-linear analysis, with any diagnostic aids such as the examination of membership lists, is here christened the 'PIE-SLICE' technique. The derivation of this name is that (amongst other things) we are slicing the total PIE (pottery information equivalent) of a site in various ways — by combination of context, form and fabric, for example. It also has echoes of the pie-chart, which may reassure the more naïve user.

We comment that real life may not be as complex as the theory. Firstly, although in theory the tables used with sub-hierarchies A, B and C are different, they may well be the same, or very similar, in practice. The imposition of archaeological constraints may restrict the possibilities so much that the tables end up the same, or almost so. Secondly, two sub-hierarchies may have the same accepted model, if at level 3 or higher.

A more ambitious approach would be to work directly with the three-way table, rather than look at each two-way marginal table in turn, as we have done. We would then have three sets of distances — between rows, columns and 'layers' — with each distance formed from a matrix-profile instead of a vector-profile as above. We have not pursued this approach because:

1. it seems to be too complicated to program reasonably,
2. it seems likely to be potentially unstable, as the individual values in the three-way table are often very small.

## Bibliography

BISHOP, Y. M. M., S. E. FIENBERG, & P. W. HOLLAND 1975. *Discrete Multivariate Analysis*. MIT Press, Cambridge, Massachusetts.

CRADDOCK, J. M. & C. R. FLOOD 1970. "The distribution of the chi-squared statistic in small contingency tables", *Applied Statistics*, 19: 173–81.

ORTON, C. R. & P. A. TYERS 1989. "Error structures of ceramic assemblages", *in* Rahtz, S. P. Q. & Richards, J. D., (eds.), *Computer Applications and Quantitative Methods in Archaeology 1989*, International Series 548, pp. 275–285. British Archaeological Reports, Oxford.

ORTON, C. R. & P. A. TYERS this volume. "A technique for reducing the size of sparse contingency tables", *in* Lockyear, K. & Rahtz, S. P. Q., (eds.), *Computer Applications and Quantitative Methods in Archaeology 1990*.