This is the manuscript of my article

J. W M. Peterson (2007) "Random orientation of Roman camps", *Oxford Journal of Archaeology,* 26.1, 103 -108.

as accepted for publication, and reproduced with the permission of the Journal and Blackwell Publishing.

Alan Richardson, the author of the article I criticised, offered some comments in response, also printed in that issue of the OJA (p108). They conclude with the suggestion that I should have made reference to another work by himself and M.J.Ferrar, entitled *The Roman Survey of Britain*, Oxford (2003). I was aware of this work but considered it to be misleading and so not worth mentioning.

J. W. M. PETERSON

RANDOM ORIENTATION OF ROMAN CAMPS

Summary. *An article on the orientation of Roman camps and forts, published in this journal, based its conclusions on faulty statistical analysis. Its only verifiable claim is that the surveyors of Roman camps seem to have preferred orientations within ten degrees of the meridian. Its other claims are not supported. The supposed anomalies of orientation do not need to be explained.*

INTRODUCTION

This is a critique of a recent contribution to these pages, concerning the orientation of Roman camps and forts (Richardson 2005). In earlier articles, Richardson has proposed that centuriation existed in the Inglewood Forest of Cumbria (Richardson 1982 1986 1999). This hypothesis, although not conclusively proved, is credible. Examination of known Roman features in the area (Peterson 2001) suggests that the hypothetical land division may indeed be the partial expression of a large multipurpose land survey.

Richardson's recent article claims that the orientation to north of a camp or fort is often the result of a planned relationship. This idea is attractive. If true, it could support the reality of the Inglewood centuriation. It would also support a general theory of Roman oblique planning. This theory derives from my own study of relative angles in Roman survey grids (Peterson 1988 1992). Simplified in the extreme, it suggests that roads

within the grids were often planned by a "join the dots" approach, where the intersections of the grid are the dots. This naturally results in lines whose slopes, relative to the grid, can be expressed by simple ratios, such as 1:1, 1:2, 2:3, etc. As Richardson says, these are the tangents of the relative angles. It is the presence of angles that have tangents like this - the ratio of small whole numbers - which could be seen as an indication that this sort of planning had taken place.

Statistical analysis is extremely important to Richardson's argument. His hypothesis depends on establishing the fact that some particular orientations were chosen in a non-random way. This presents a problem because, even at first glance, the analysis seems flawed. I have no objection in principle to many of the ideas in the article, but have to ask if its claims are really supported by the observations.

THE CHI SQUARE TEST - MINIMUM BIN SIZES

The first part of Richardson's analysis is intended to show that some orientations of a set of camps occur more frequently than would be expected if each were independently chosen at random. It is claimed (Richardson 2005, 417) that a statistical test, chi square, gives a clear and very positive result.

Many of us feel uneasy when dealing with small samples. It is therefore natural to look carefully at a claim based on the common orientation of only six camps. This leads to detailed consideration of the information presented in Richardson's Table 2.

As Richardson says, because of the symmetry of the camps and forts, and because their orientations with respect to the meridian can be either east or west, relative orientations cover a range of 45 degrees. Since he measures the angles to a precision of one degree, the data in his Table 2 are the values of a histogram, with 67 items assigned to 45 bins, according to their orientation. In this case the number of items in each bin varies from zero (sixteen bins) to six (one bin). The question is whether or not this distribution is abnormal.

The calculations can be repeated. The observed values may be entered in a spreadsheet column containing 16 zeros, 12 ones, 7 twos, 4 threes, 2 fours, 3 fives and a six. An adjacent column of expected values may be created, each cell containing 45 copies of 67/45. This last figure is (formally) the most likely value of the bin contents if 67 items are assigned at random to 45 bins. To three decimal places, this is 1.489.

The chi square calculation is straightforward. For each pair of observed and expected values, the squared difference between the two is divided by the expected value. These results are then summed. In this case, the result is 80.09 (to two decimal places).

The Excel function CHIDIST, using this figure for chi square, produces a value of 0.000718. On the face of it, this seems be the probability of seeing this distribution at random. Richardson, who presumably used similar (if not the same) software to obtain similar results, consequently says that the distribution is "highly abnormal" because "p < 0.001".

With the best will in the world, this statement cannot be supported. In this case, the value generated by the chi square test is not evidence that the distribution is abnormal.

In order to see why this is so, it is vitally important to appreciate that the chi square statistic is calculated by a mechanical procedure, deriving a value from the difference between observed and expected values in each bin. It calculates how much the observed and expected bin contents, as a whole, differ from each other. However, its output cannot validly be used if the characteristics of the input data are unwisely chosen.

One known cause of an invalid result is a choice of bins whose expected values are too small. A Google search on "chi square expected valid" produces 1,600,000 results. Many give advice on expected values required to obtain valid results. Frequently they say that five should be the minimum expected in any bin, although some sources, such as the Mathematics Learning Support Centre, Loughborough University, are slightly more lenient, allowing that up to 20% of the expected values may be less than five.

Hence a test with expected values of 1.489 in all bins cannot produce a valid result. The supposed odds of less than 1:1,000 are not reliable.

To a mathematician, and perhaps to an archaeologist, the immediate question is what the odds really are. If the 67 orientations are measured to a degree, we know that a chi square test cannot validly answer this question. Nor does there appear to be a simple analytical method or

computation that will do so. However, a Monte Carlo simulation can quickly provide a rough answer. This is naïve approach, but it may provide a credible result.

Imagine that we have 45 buckets and 67 balls. We throw all the balls at the buckets at random, so that each ball has an equal chance of landing in any bucket. We then count the number of balls in each bucket, retrieve the balls and have another go. Having done this many times, we begin to see how frequently, in all our trials, we will find a bucket with six or more balls in it.

Now, instead of imaginary balls and buckets, you may use a spreadsheet. Create a table of 45 columns (one for each bucket) and 67 rows (the balls). Place above the columns a row of numbers, 1 to 45; these are the column numbers. Place beside the rows a column of 67 cells, each containing one of the numbers 1 to 45 chosen at random; these are the destined bucket number for each ball. For each cell in the table, set it to one if its column number and destined bucket number are the same, otherwise set it to zero. The sum of each column then gives the number of balls ending up in each bucket on this trial. Record the largest of these values and repeat the process with a different assignment of balls to buckets.

A small experiment was conducted in this way - readers might care to repeat it with their own spreadsheet. The simulation was run until, on 30 occasions, a bucket had been observed containing at least six balls. In three such sequences this required 189, 166 and 171 runs respectively. Over the

first 150 runs (an arbitrary figure) the number of times at least six balls turned up was 23, 27 and 28. This is a frequency of about 17%, give or take a few percentage points. A longer simulation could be conducted to obtain a more accurate estimate, but this does not seem worth the effort. This result is enough to show that we should expect the odds of to be about 1:6. Since this is very different to 1:1,000, it is proof, if proof were needed, that one cause of unreliability of this chi square test is a choice of bucket size for which expected values are too small. It also shows that Richardson's distribution of individual angles cannot be considered to be significantly abnormal.

Before proceeding to describe another possible source of error, it might be helpful to explore how Richardson's data, in its raw form, could have been handled in order to make the chi square test valid.

When expected values in bins are too small, the standard advice is to group data categories. In this case all expected values are the same. The simplest way to group the data is in ten ranges of four degrees and one of five. The bin sizes do not have to be the same, but naturally the expected value in a larger bin will be greater than in the others. The grouped data are shown in Table 1.

[Insert Table 1 about here]

The resultant probability is not significant. The distribution cannot be regarded as significantly abnormal. No meaning can be ascribed to it.

On a more positive note, one of Richardson's observations can be supported. It may be meaningful that 26 of the orientations fall in the range zero to nine degrees. A chi square test, comparing the number of orientations within this range to the number at other orientations, indicates statistical significance. The resulting probability is 0.0011, which is significant ($p < 0.002$). When the data are considered in this way, the odds are longer than 1:500 that it is a random distribution. Using a valid test, there is a perceptible preference for orientations around the meridian, within a range of 20 degrees.

APPLICATION OF THE CHI SQUARE TEST - INDEPENDENCE OF OBSERVATIONS

Another condition for a valid application of a statistical test is that the observations are independent.

Unlike a breach of the rule on minimum bin size, lack of independence is debatable. It may be hard to see because of more or less hidden connections between observations. They require close scrutiny in order to rule out the more obvious dependencies. An examination of one of Richardson's sources illustrates this point.

Ardoch camps 1, 3, 4 and 6 (St Joseph 1977, 136) are among the six that, according to Richardson's appendix, Table A1, have a common orientation. He measures their bearing to the meridian as 42 or 132 degrees. Their axes of symmetry, if assumed to be at right angles, are thus parallel to each other.

The sceptical reader might care to question whether or not camp 1 deserves to be included with the others but, because the purpose of this article is solely to examine the use of methods, these data will be taken at face value.

Dependency seems probable between the orientations of camps 3 and 4. It is hard to believe that the surveyors of one camp had no knowledge of the layout of the other. Otherwise, why do the parallel sides line up? Erring on the side of caution, we should be prepared to accept that one was derived from the other, and that their common orientation was chosen only once.

Furthermore, we cannot be certain that the orientation of the other camps was chosen independently from that of camps 3 and 4. They are only 500m apart, so there may have been locally some structure or boundary, not discernible in the aerial photography, which determined the orientation of all of them.

Therefore, to increase the chance that the data are truly independent, it would be wise to represent commonly orientated structures at a single site by a single data item. The effect of this on the data in Richardson's Table A1 would be to reduce, to one, the two fort orientations at two other sites, Walton (both 74 degrees to the meridian) and Chew Green (both 122 degrees). With four items reduced to one at Ardoch, this would reduce the total data set to 62 items, and the largest bin contents to five.

Further Monte Carlo simulation with the reduced data set shows that 200 trials were required in order to observe a bucket containing five or more balls on 90 occasions. This is an expectation of 45%. Hence, if 62 orientations are chosen at random, the toss of a coin could decide quite well whether or not, on any particular trial, we see at least five the same. Five camps, out of 62, share a common orientation, but this observation is certainly not abnormal.

DISCUSSION

Examination of the way in which these data have been processed reveals some disturbing shortcomings. In at least one instance, no consideration seems to have been given to the issue of data independence. It may be debatable, but it is not debated.

Much more disturbing is the execution of the chi square test. A test conducted in this way could not, and does not, produce a valid result. It is indubitably wrong to insist that the null hypothesis be rejected (Richardson 2005, 420). This has disastrous consequences for nearly all of the article's discussion and conclusions. They have no foundation.

One observation in the article may still have meaning. Let us assume that the set of camp orientations is a truly arbitrary sample of the orientations of all camps, and let us also ignore possible dependencies. In that case we have a set of orientations that seems to have a significant preference for an approximately polar or equatorial direction. It is, however, an indecisive preference. Since it has an uncertainty of up to ten degrees either way, it

could be determined adequately by knowledge of where the sun was at the time of a midday meal.

The other results are meaningless. There is no significant preference for any individual angle within a range of 45 degrees. Hence, there is no need for explanation. There is nothing to explain.

How might such problems be avoided?

Authors have the primary responsibility for the accuracy of their claims, but they may need help. Advice should be sought.

Software could also be more helpful to its users. The Excel function CHITEST will blithely calculate a chi square value from two columns of data, even if the expected values are such that the outcome of the test is meaningless. It would have taken very little extra software programming to check the values against well-recognised advice, and to give an appropriate warning if necessary.

CONCLUSIONS

An article published in this journal reached conclusions based upon faulty statistical analysis. Its only verifiable claim is that the surveyors of Roman camps seem to have preferred orientations within ten degrees of the meridian. The remainder of the argument has no foundation.

Authors who wish to analyse their data should seek advice on appropriate methods. Computerised statistical software should more frequently alert users in situations where results are likely to be untrustworthy.

School of Computing Sciences

University of East Anglia

Norwich NR4 7TJ

UK

Email: jwmp@cmp.uea.ac.uk

REFERENCES

PETERSON, J. W. M. 1988. Information systems and the interpretation of Roman cadastres. In Rahtz, S. P. Q. (ed.) *Computer and Quantitative Methods in Archaeology: CAA 88* (Oxford, BAR Int. Ser. S446), 133-149.

PETERSON, J. W. M. 1992. Trigonometry in Roman cadastres. In Guillaumin, J.-Y. (ed.), *Mathématiques dans l'Antiquité* (Université de St-Étienne, Centre Jean-Palerne: Mémoires 11), 185-203.

PETERSON, J. W. M. 2001. Design and Performance of the Varatioscope. In Stancic, Z. and Veljanovski, T. (ed.), *Computing Archaeology for Understanding the Past: CAA 2000* (Oxford, Archeopress), 269-272.

RICHARDSON, A. 1982. Evidence for Centuriation in the Inglewood Forest. *Transactions of the Cumberland and Westmorland Antiquarian & Archaeological Society* 82, 67-71.

RICHARDSON, A. 1986. Further Evidence of Centuriation in Cumbria. *Transactions of the Cumberland and Westmorland Antiquarian & Archaeological Society* 86, 71-78.

RICHARDSON, A. 1999. Centuriation South of Hadrian's Wall in Cumberland. *Matterdale Historical and Archaeological Society Year Book and Transactions* 5, 7-14.

RICHARDSON, A. 2005. The orientation of Roman camps and forts. *Oxford Journal of Archaeology* 24(4), 415-426.

ST JOSEPH, J. K. 1977. Air Reconnaisance in Britain, 1973-1976. *Journal of Roman Studies* 67, 125-61.

TABLE 1

FREQUENCY OF ORIENTATION ANGLES

(Expected frequencies: 4 degree range - 5.96;
5 degree range - 7.44)

Angles (degrees)	Frequency
0 - 3	10
4 - 7	8
8 - 11	10
12 - 15	2
16 - 19	10
20 - 23	6
24 - 27	4
28 - 31	3
32 - 35	6
36 - 39	2
40 - 44	6

Chi Sq = 16.6, df = 10, $p < 0.1$, $p = .084$