A Statistical Approach to Test Pitting

Rich Wareham
Signal Processing and Communications Laboratory
Department of Engineering, University of Cambridge
Email: rjw57@cam.ac.uk

***Abstract*—We develop a simple threshold-based test for determining the significance of one test pit when compared to others within some shared context, for example a single village. We do this by developing a statistical model for test pit evolution and then use the model to provide a data-driven test for significance based on Welch's t test[1]. For our purposes, 'significant' is taken to mean 'surprising given the wider context'. Such significant test pits can be used to localise settlement or other processes leading to larger-scale pottery deposition.**

I. Introduction

In this text we consider a statistical approach to analysing test pit data in which we assign a number indicating the significance, in a statistical sense, of a given site given the wider context of the area. We do this by comparing sherd counts which are partitioned into different eras.

Conceptually we wish to find some statistic which reflects our prior belief that a test pit which has a larger proportion of, for example, Roman pottery in it in comparison to other test pits in the area is in some way more 'surprising', i.e. has a lower probability, than test pits whose pottery proportions match the site as a whole. We aim that this statistic be useful for 'calling attention' to interesting test pits.

Throughout this document we shall consider a single test pit within a collection of pits on the same site and use the count of pottery as the measured value. The mathematics works for both pottery weight or sherd counts, although sherd counts are more convenient, and may trivially be extended to consider the probability of entire sites within wider contexts of settlement by considering an individual site as a single large test pit. Indeed, such an approach is likely, given the greater quantities of data, to lead to firmer conclusions at the expense of coarser localisation.

For the purposes of illustration, we will consider pottery to fall into the categories Roman, Early Medieval, Late Medieval, High Medieval. This categorisation is purely conventional and the technique below will work with any pottery categorisation assuming that no one sherd may fall into more than one category and that all sherd being considered have an associated category.

II. Notation

Assume that there are N test pits in the site. For the i-th test pit we denote the count of Roman era pottery sherds as R_i, the count of Early Medieval pottery as E_i, Late Medieval as L_i and High Medieval as H_i. The total count of pottery sherds is denoted as T_i and we make the assumption that all pottery in the pit comes from one of our designated eras. That is to say that we assume

$$T_i = R_i + E_i + L_i + H_i.$$

Given these individual test pit counts we can derive the *total* count. For example the total count of Roman pottery, R, is given by

$$R = R_1 + R_2 + \cdots + R_N \equiv \sum_{i=1}^{N} R_i$$

and similar expressions can be written down for the total amounts of Early, Late and High Medieval, E, L and H respectively and also the total count of pottery over all pits, $T = R + E + L + H$.

Sometimes, for convenience's sake, we refer to the *normalised* counts, for example $R'_i \equiv R_i/T_i$ which can be interpreted as the *proportion* of, in this instance, Roman pottery within the pit.

III. Welch's t test

If we let $R_{\setminus i}$ and $T_{\setminus i}$ be the Roman era and total counts *excluding* test pit i then we note that we have two completely independent estimators for the proportion of Roman pottery on the site:

$$R'_i = \frac{R_i}{T_i}, \quad \text{and,} \quad R'_{\setminus i} = \frac{R_{\setminus i}}{T_{\setminus i}}.$$

If test pit i is unsurprising, we expect R'_i to approximately equal $R'_{\setminus i}$. Welch's t test[1] can be used to determine if any differences are statistically significant. To make use of this test we interpret R'_i as the mean of T_i samples. In the case of sherd counts, these are R_i ones and $T_i - R_i$ zeros. For sherd weights, we need to know the total number of sherds and their individual weights to perform the same analysis. This information is usually unavailable and so we proceed assuming we're dealing with sherd counts. Specifically, for each test pit, i, one computes the statistic

$$t_i^{(R)} = \frac{A_R^{(i)}}{\sqrt{B_R^{(i)}}}$$

where

$$A_R^{(i)} = \frac{R_i}{T_i} - \frac{R_{\setminus i}}{T_{\setminus i}}, \quad \text{and,} \quad B_R^{(i)} = \frac{\sigma_i^2}{T_i} + \frac{\sigma_{\setminus i}^2}{T_{\setminus i}},$$

and σ_i^2 and $\sigma_{\setminus i}^2$ are the sample variances.

For sherd weights, these variances would have to be calculated by measuring the weight of each individual sherd but for sherd counts we can make use of a trick: we know the samples for test pit i consist of R_i ones and $T_i - R_i$ zeros,

$$\sigma_i^2 = \frac{1}{T_i - 1}\left[R_i\left(1 - \frac{R_i}{T_i}\right)^2 + (T_i - R_i)\left(-\frac{R_i}{T_i}\right)^2\right]$$

by simple application of the formula for sample variance. Multiplying out, we obtain:

$$\begin{aligned}\sigma_i^2 &= \frac{1}{T_i - 1}\left[R_i - \frac{2R_i^2}{T_i} + \frac{R_i^3}{T_i^2} + \frac{T_i R_i^2}{T_i^2} - \frac{R_i^3}{T_i^2}\right] \\ &= \frac{1}{T_i - 1}\left[R_i - \frac{R_i^2}{T_i}\right] = \frac{T_i R_i - R_i^2}{T_i(T_i - 1)}.\end{aligned}$$

One can similarly obtain a relation for $\sigma_{\setminus i}^2$ and hence,

$$B_R^{(i)} = \frac{T_i R_i - R_i^2}{T_i^2(T_i - 1)} + \frac{T_{\setminus i} R_{\setminus i} - R_{\setminus i}^2}{T_{\setminus i}^2(T_{\setminus i} - 1)}.$$

There exists a standard formula[1] for the degrees of freedom, $\nu_i^{(R)}$, from which we can obtain a p-value for the null hypothesis via the Student's t-distribution which is parameterised by $\nu_i^{(R)}$. For our purposes, however, it is usually sufficient to make use of the rule of thumb that

$$\left|\frac{A_R^{(i)}}{\sqrt{B_R^{(i)}}}\right| \geq 3$$

indicates a statistically significant result. (For degrees of freedom from 1 to $+\infty$ this rule is reasonable.) For completeness, the relation for $\nu_i^{(R)}$ is

$$\nu_i^{(R)} = \frac{\left(B_R^{(i)}\right)^2}{C_R^{(i)}}$$

where

$$C_R^{(i)} = \frac{\sigma_i^4}{T_i^2(1 - T_i)} + \frac{\sigma_{\setminus i}^4}{T_{\setminus i}^2(1 - T_{\setminus i})}.$$

IV. Conclusion

In the section above we obtained a statistic which may be used to evaluate the significance of one test pit in comparison to others within a similar location. Specifically, for the example of Roman pottery, the value

$$t_i^{(R)} = \frac{R_i \cdot T_i^{-1} - R_{\setminus i} \cdot T_{\setminus i}^{-1}}{\sqrt{\frac{T_i R_i - R_i^2}{T_i^2(T_i - 1)} + \frac{T_{\setminus i} R_{\setminus i} - R_{\setminus i}^2}{T_{\setminus i}^2(T_{\setminus i} - 1)}}}$$

may be computed for the i-th test pit and can be interpreted as a number of standard deviations away from the null hypothesis. Consequently, a value of $|t_i^{(R)}| \geq 3$ for test pit i can be used to conclude that the proportion of Roman pottery in test pit i differs from neighbouring test pits in a statistically significant way. The sign of $t_i^{(R)}$ indicates whether this differences is in the form of a surprisingly little (-ve) or a surprisingly large (+ve) amount of pottery. Similar expressions can be formed for the other eras.

It should be noted that this statistic is merely a value which can be used to directly compare test pits based on the assumptions stated above. Specifically, we have assumed that test pits which contain a relatively large amount of pottery from one era or a relatively large amount of pottery in general are more indicative of the "true" pottery distribution and that such indicative test pits which vary greatly from the rest of the site are significant. The aim of the statistic is to flag those test pits worthy of closer inspection. Any conclusions which may be drawn from those pits are better left to an archæologist directly inspecting the nature of the sherds.

References

[1] Welch, B. L. (1947). *The generalization of "Student's" problem when several different population variances are involved*. Biometrika **34** (1–2): 28–35. doi:10.1093/biomet/34.1-2.28. MR 1.