NEW ZEALAND ARCHAEOLOGICAL ASSOCIATION

NEW ZEALAND ARCHAEOLOGICAL ASSOCIATION NEWSLETTER

Multi-sampling and Absolute Dating Methods:

A Problem of Statistical Combination for Archaeologists[1]

B. F. Leach
Department of Anthropology
University of Otago

Summary: A method for pooling absolute dates is described with examples of its application

When facing the problem of assessing the absolute age of a prehistoric deposit, archaeologists commonly encounter circumstances which involve multi-sampling procedures. For example, it has become a commonplace to select two or more charcoal samples from the same archaeological provenance and submit these to one or more C^{14} laboratories for individual age determinations. It is most unlikely that the results returned will be identical. Other possibilities also arise, such as several age assessments of identical material by different methods (for example collagen and carbonate dating of bones), and by application of different methods to different materials from identical archaeological contexts (such as the C^{14} method with charcoal, and hydration rim dating of obsidian).

While these multi-sampling procedures are widely employed there has been little agreement as to any standard method by which several age determinations may be combined to yield an overall 'best estimate' of age for the horizon from which the samples originally derived. Indeed, it seems there may be some confusion as to the statistical meaning of absolute age statements which might be received from a dating laboratory. It is commonly stated, for example, that the 'statistical error' associated with C^{14} dates is the standard deviation (σ). This is certainly true, but what is possibly not commonly realised is that the referent variable of σ is not x (the original observations in the laboratory) but $\bar{x}$ (the mean observation or equivalent age). The statistic which is reported is actually $\sigma x/\sqrt{N}$ or the standard error of $\bar{x}$. This is equivalent to $\sigma\bar{x}$, hence the confusion. There is a subtle, though important, difference between the two terms, and in some ways it would be wiser to refer to the error as the 'standard error of the age' to obviate any misconceptions in future. It should be noted that the important differ-

1 I am grateful to Dr. Rafter of the D.S.I.R. for suggesting the statistical formulae used in this paper.

ence between the standard error of $\bar{x}$ and the standard deviation of x consists in the fact that if the number of observations is increased (perhaps by submitting several charcoal samples) the former statistic may lessen considerably, while the latter will probably remain the same. That is to say, the standard deviation of the observations will be unchanged, while the estimation of the population mean (the 'true' sample age) becomes putatively more accurate.

Under normal circumstances, when only one absolute age is available, there is no real problem. On the other hand, if multi-sampling is involved, estimating overall age is by no means straightforward. In the past, archaeologists have followed any number of different 'rules of thumb'. In many situations, especially in New Zealand, where archaeological ages seldom exceed 1000 years B.P., the material difference between following one rule of thumb and another will not be great. At the same time, a standard method for pooling such results is clearly desirable, and in certain cases very necessary. Such a case would be with the estimations of age by measurement of hydration rims on obsidian. A single archaeological context may be dated by perhaps as many as twenty different pieces of obsidian, each of which is measured several times and has its own standard error. The problem of pooling results is unavoidable and directly comparable to pooling several C^{14} ages.

In the field of absolute dating this problem was first faced in the determination of an acceptable value for the half-life of C^{14}. As is well known, the figure accepted internationally was derived from carefully controlled experimental conditions involving three different laboratories. The several results for the C^{14} half life are as follows:

5580 yrs ± 45
5589 yrs ± 75
5513 yrs ±165

These were statistically combined to yield a best estimate commonly referred to as the 'Libby' half life:

5568 yrs ± 30

It has been difficult to discover precisely what method was used to derive this 'best estimate', but it must have been closely similar to that used widely by statisticians with similar problems, namely:

Original ages (before present)

$$A_1 \pm E_1,\ A_2 \pm E_2,\ A_3 \pm E_3,\ \ldots\ A_n \pm E_n$$

Best estimate of age = A

$$A = \frac{\Sigma \frac{A_i}{E_i}}{\Sigma \frac{1}{E_i}}$$

Best estimate of standard error = E

$$E = \sqrt{\frac{\Sigma (A_1 - A)^2}{n(n-1)}}$$

An alternative formulation is to use the values of $(E_i)^2$ in the determination of A rather than as stated above. The difference in method is academic and seldom produces very different results. By using these formulae, the three radiocarbon determinations quoted above yield a best estimate for the C^{14} half life of:

(a) (based on standard errors) 5573 yrs ±26

(b) (based on standard errors squared) 5579 yrs ±27

It will be noted that this departs slightly from the internationally accepted value, and presumably results from the use of a somewhat different method.

The method suggested above possesses two noteworthy features. Firstly, the derived best estimate of age takes due account of varying standard errors associated with the initial determinations, and weights the overall assessment towards those with the lowest errors. Secondly, the derived best estimate of the standard error is a far more satisfactory statement of the probability range of overall age than any rule of thumb method. Most archaeologists would intuitively feel that two closely similar C^{14} dates provide a rather more secure overall date; the method suggested results in a reduced standard error in such cases.

The following two examples will indicate the type of results which can be achieved with the method.

Example 1:

Three very small charcoal samples recovered from a stone walled enclosure-garden in Palliser Bay. While they each derive from different stratigraphical provenances, it is not believed that deposition involved a lengthy period of time.

R2850/10	1442 A.D.	±79
R2850/8	1562 A.D.	±79
R2850/9	1608 A.D.	±78

The combined results are as follows (N.B. a is based on standard errors, and b on standard errors^2):

(a)	1538 A.D.	±49
(b)	1538 A.D.	±49

Example 2:

Two charcoal samples from the same layer in the Oturehua quarry site in central Otago yielded:

R2054/2	1053 A.D.	±27
R2233/2	1023 A.D.	±82

The combined results are:

(a)	1046 A.D.	±17
(b)	1050 A.D.	±19

As can be seen from these examples, little difference results from using E_n or E_n^2 in the formula, and perhaps the simpler non-squared version could be adhered to for the sake of uniformity. In conclusion, it should be noted that this method can easily result in a false sense of secutiry as to the age of an archaeological horizon. The derived best estimates are applicable to the age of the samples, and rather more tenuously to the age of the prehistoric context. Uncertainties surrounding for example the original position in the tree of a charcoal sample, and the true half-life of C^{14}, to mention but two, combine to make an error of ±17 years (as in Example 2 above) a rather unrealistic appraisal of contextual age. Used with caution however, the method should be acceptable to mathematicians and archaeologists alike.