Ancient TL

www.ancienttl.org · ISSN: 2693-0935

Galbraith, R., 2003. *A simple homogeneity test for estimates of dose obtained using OSL.* Ancient TL 21(2): 75-77. https://doi.org/10.26034/la.atl.2003.362

A simple homogeneity test for estimates of dose obtained using OSL

Rex Galbraith
Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, U.K.

(Received 2 January 2004; in final form 19 February 2004)

In optical dating (and in other dating methods) it is often required to compare age estimates, or equivalent dose estimates, for a number of single grains or aliquots. Typically each estimate has a different and (hopefully) known precision, which needs to be taken into account. A useful graphical method in this situation is the radial plot (Galbraith, 1988; Galbraith *et al.,* 1999) in which standardised estimates are plotted against the reciprocals of their standard errors. One feature of this plot is that a set of estimates that agree with each other, within error, will scatter homoscedastically, with unit standard deviation, about a line through the origin. This gives a visual assessment of which estimates are consistent with a common age or dose.

Often such a visual assessment will be sufficient, but sometimes it may be useful to assess more formally whether several estimates are consistent with a common value. This note points out that there are standard statistical tests available for this purpose.

Homogeneity test

Suppose that we have n independent estimates $z_1, z_2, ..., z_n$ with standard errors $\sigma_1, \sigma_2, ..., \sigma_n$. Let μ_i denote the expected value of z_i, so we can regard z_i as an observation from a distribution with mean μ_i and standard deviation σ_i. We wish to test the null hypothesis that $\mu_i = \mu$, a common (but unknown) value for each i.

There is a standard test based on the assumption that z_i is from a Normal distribution. In this case, the maximum likelihood estimate of μ (under the null hypothesis) is the weighted mean

$$\hat{\mu} = \frac{\sum_{i=1}^{n} w_i z_i}{\sum_{i=1}^{n} w_i} \quad \text{where} \quad w_i = \frac{1}{\sigma_i^2}$$

and the homogeneity test statistic is

$$G = \sum_{i=1}^{n} w_i (z_i - \hat{\mu})^2 .$$

If the μ_i are all equal, then G will be from a χ^2 distribution with $n-1$ degrees of freedom. If the μ_i differ then G will tend to be larger.

So the test is to calculate G and hence calculate the "P-value", which is the probability that a random value from the χ^2 distribution with $n-1$ degrees of freedom is greater than G. A significantly small P-value, with the usual conventions (e.g., less than 0.05 or 0.01), provides evidence that the μ_i are not all equal.

It is worth noting immediately that a significant P-value can also arise if the standard errors σ_i are not correct, particularly if they under-estimate of the true standard errors. This possibility should be borne in mind in practice. To put it another way, one is really assessing whether the observed variation in z_i is consistent with what would be expected from the given σ_i alone.

Relation with radial plots

The above test can be interpreted in terms of a radial plot, where $y_i = z_i/\sigma_i$ is plotted against $x_i = 1/\sigma_i$. The estimate of μ can be written equivalently as

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

which is the slope of the ordinary least squares regression line through the origin; and G can be written as

$$G = \sum_{i=1}^{n} (y_i - \hat{\mu} x_i)^2$$

which is the sum of squared residuals about the regression line. In effect, the test is assessing whether the residual standard deviation about the regression line through the origin is consistent with 1. In practice, in a radial plot one normally uses $y_i = (z_i - z_0)/\sigma_i$ for some convenient z_0, but the above interpretation still applies.

Application to OSL doses

Many optical dating methods use a single aliquot regenerative dose (SAR) protocol, which produces a dose estimate and its standard error for each grain or aliquot in a sample. Suppose we wish to assess whether these estimates are consistent with a common value. There is more than one way to apply the above test. For example, we could let z_i be the dose estimate (in Gy), or we could let z_i be the natural *log* of the dose estimate. Because the test assumes that z_i is from a Normal distribution, the latter choice may often be preferred. In this case, σ_i would be the standard error of the log dose — which is effectively the *relative* standard error of the dose. Note furthermore that the standard error of a dose estimate generally increases with the size of the dose, whereas the relative standard error does not (to the same extent, at least). This is another reason why using the log doses may be preferred.

For illustration, consider the data in Table 1, taken from Galbraith *et al.* (1999). These are palaeodose estimates and standard errors (in Gy), and the corresponding log palaeodose estimates and their standard errors, for seven single grains of quartz.

Grain number	Palaeodose (Gy)			Log palaeodose	
	estimate	s.e.		estimate	s.e.
19	30.1	4.8		3.4055	0.1607
22	53.8	7.1		3.9857	0.1314
25	54.3	6.8		3.9943	0.1253
50	29.0	4.3		3.3663	0.1494
99	47.6	5.2		3.8630	0.1087
105	44.2	5.9		3.7887	0.1330
107	43.1	3.0		3.7627	0.0702

Table 1.
Data for example calculation

In this example it is more reasonable to apply the test to the log palaeodose estimates. Indeed, the standard errors of these were obtained directly from the OSL photon counts, while the standard errors of the palaeodoses were derived from them — e.g., for grain 19, as $30.1 \times 0.1607 = 4.8$.

The reader may verify that, from the above formulae, $\hat{\mu} = 3.7737$ and $G = 19.10$, with 6 degrees of freedom. Hence the P-value is approximately 0.004 (the probability that a value from $\chi^2(6)$ is greater than 19.10), which is strong evidence against the null hypothesis. On this basis there is strong evidence that the estimated doses are not consistent with a common value, within the given errors. As noted above, this may be because the true doses vary or because the standard errors have been under-estimated (or both).

Figure 1 shows a radial plot of the data in Table 1 which suggests visually that the estimates are slightly over-dispersed — for example, two of the seven points scatter outside the ± 2 standard error band shown. This plot uses $z_0 = \hat{\mu} = 3.7737$, so the horizontal radius corresponds to $e^{3.7737} = 43.5$ Gy.

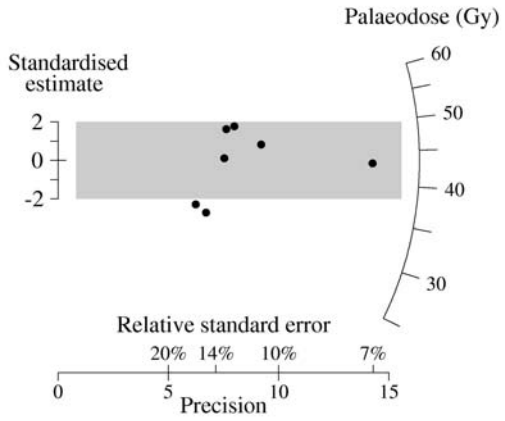

Figure 1.
Radial plot (log scale) of the palaeodose estimates in Table 1.

Galbraith *et al.* (1999) used these data to illustrate the calculations for the central age model. The over-dispersion parameter σ was estimated to be 0.1724, with standard error 0.0694, suggesting, in particular, that σ is greater than zero (rather than equal to zero). That is, the dose estimates vary by more than their standard errors would imply, in agreement with the above test. The central age model, in addition, gives an estimate of the amount of over-dispersion (albeit rather an imprecise one in this example with only seven grains).

Further remarks

Homogeneity tests, such as that above, have been found useful in other contexts — for example in fission track dating and in "meta analysis" of medical trials. When over-dispersion is present, different types of question may then arise as to its cause and effect, which may invite different types of subsequent

analysis. A good modern reference to the above test is Armitage *et al,* (2002), pages 216 and 643. This is a new edition of a classic book aimed at medical researchers, but it is also an excellent general reference for modern statistical methods.

The above test has quite low power, particularly when n is large and the μ_i do not vary greatly. That is, moderate heterogeneity of the μ_i may not produce a significantly large G. This may not be a bad thing in the present context. As always, it should be remembered that data may be consistent with the null hypothesis and at the same time be consistent with other hypotheses.

Finally, the fact that standard errors of dose estimates tend to increase with the size of dose, has other implications. For example, when plotting a number of dose estimates in a histogram, the larger ones will tend to scatter more. This is one reason why such histograms are often positively skewed.

I thank Geoff Duller and Bert Roberts for useful comments.

References

Armitage, P., Berry, G., and Mathews J.N.S. (2002) *Statistical Methods in Medical Research*, 4th Edition, Blackwell, 2002.

Galbraith, R.F., (1988) Graphical display of estimates having differing standard errors. *Technometrics,* 30**,** 271–281.

Galbraith, R.F., Roberts, R.G., Laslett, G.M., Yoshida, H. and Olley, J.M. (1999) Optical dating of single and multiple grains of quartz from Jinmium rock shelter, Northern Australia: Part I, Experimental design and statistical models. *Archaeometry,* 41**,** 338–364.

Reviewer

Geoff Duller