Terrestrial mass extinctions and galactic plane crossings

IN their attempt to build a case for a "galactic plane crossing" explanation for terrestrial extinction, Rampino and Stothers[1] present a statistical argument that is seriously misleading. The crux of the empirical component of their argument is a claimed agreement between two columns of dates (their Table 1):

Table 1 Dates from ref. 1

Mass extinctions (Myr BP)	Galactic plane crossings (Myr BP)
11	0
37	31
66	64
91	100
144	135
176	166
193	197
217	227
245	259

They begin by calculating the correlation coefficient of these two columns to be $r=0.996$. It is a mistake to attribute any meaning to this number: the correlation of any two monotonically increasing sequences is bound to be high. This reflects only the high serial correlation of the separate series, and in no way indicates any connection between the two series other than a common monotonicity. For example, the correlation between the list of mass extinction dates and the first nine prime numbers (2, 3, 5, 7, 11, 13, 17, 19, 23) is $r=0.986$.

Rampino and Stothers then go on to compare the gaps in the two columns by Student's t-test for matched pairs. There are two serious difficulties in this; the point will be clearer as we introduce some notation. Let $X_1=11,\ldots,X_9=245$ be the entries in the first column and $Y_1=0,\ldots,Y_9=259$ be the entries in the second column. Then Rampino and Stothers apply Student's t-test to the eight differences $Z_i=(Y_{i+1}-Y_i)-(X_{i+1}-X_i)$. The test statistic is then the average of the Zs divided by what would be an appropriate estimate of standard error of that average if the Zs were independent. But the average of the Zs reduces algebraically to $\bar{Z}=[(Y_9-Y_1)-(X_9-X_1)]/8$; it does not depend on the intermediate values at all. If the two columns spanned the same time interval, it would reduce identically to zero, regardless of the intermediate values. In other words, the Zs are highly correlated because adjacent intervals share common endpoints; the test as performed would be valid only if both series were random walks (hardly a tenable hypothesis, particularly for galactic plane crossings). Consequently, the denominator of the t-statistic grossly over estimates the standard error of the numerator and biases the test statistic towards zero. Even if their test were valid, there is a second difficulty in this aspect of their argument: they interpret a small t-value as evidence in favour of a null hypothesis, when it could as well be held to reflect a paucity of data and a test insufficiently powerful to detect a difference.

S. M. STIGLER

Department of Statistics, University of Chicago, Chicago, Illinois 60637, USA

1. Rampino, M. R. & Stothers, R. B. *Nature* 308, 709–712 (1984).

RAMPINO AND STOTHERS REPLY— Stigler regards as misleading the two statistical tests that we[1] made of the correlation between two observed time series, one containing the dates of the last nine crossings of the galactic plane by the Solar System[2], and the other consisting of the dates of the last nine major mass-extinction episodes on Earth, which we selected from the marine extinctions record published by Raup and Sepkoski[3]. Both records cover the past 260 Myr.

In our first test, we computed the correlation coefficient between the two series and obtained $r=0.996$. Stigler correctly points out that any two monotonically increasing series have a numerically high correlation. As a random test case, he correlates the first nine prime numbers with our mass-extinctions time series and obtains $r=0.986$. To make a better comparison, however, we have computed 5,000 random time series, each containing nine dates drawn from a fixed time range (0–260 Myr). These series have been correlated with our mass-extinctions time series. We find agreement with Stigler in that a statistically significant percentage of cases (13%) have $r\geqslant 0.986$.

On the other hand, only an insignificant 0.4% of the cases computed in these Monte Carlo simulations show $r\geqslant 0.996$. Furthermore, if we correlate our random time series with the galactic time series, the percentage is even lower, 0.2%. Stigler's mistake, therefore, is to regard $r=0.996$ as being close to $r=0.986$ in the case of two monotonically increasing series with nine members each.

With regard to our second test, the matched-pairs t-test applied to the time intervals, Stigler again makes an invalid criticism, because the data for any matched-pairs t-test at all can be recast and analysed in exactly the way he presents, with the same conclusion. Consider, for example, two random samples drawn from independent populations and containing eight elements each: $A_1,\ldots,A_8$ and $B_1,\ldots,B_8$. Enumerate a series of nine Xs and a series of nine Ys by computing $A_i=X_{i+1}-X_i$ with any X_1 and $B_i=Y_{i+1}-Y_i$ with any Y_1. The t-test is now straightforwardly applied to the differences $Z_i=B_i-A_i$. Because the test statistic t is proportional to the average of the Zs, we compute $\bar{Z}$ and find after some manipulation

$$\bar{Z}=[(Y_9-Y_1)-(X_9-X_1)]/8$$

which is identical to Stigler's result. Thus, $\bar{Z}$ appears not to depend at all on the intermediate values of the Xs and Ys, as Stigler noted. Stigler's result, therefore, is seen to be just a mathematical rearrangement of the data. As long as the sample elements, the As and Bs, are independent and approximately normally distributed, the t-test is valid. In our case, the sample elements are the time intervals in the time series, which are the relevant, physically independent units and are affected by random errors arising from many complex physical and accidental factors for both series[1–3]. The time intervals are, in fact, found to be approximately normally distributed. Thus, these time series can be regarded as random walks. From the point of view of testing the equality of the averages of the time intervals in the two series, it makes no difference in what order the time intervals are taken, although, if pairs are to be matched, the members of pairs must be kept together.

Our use of the matched-pairs t-test appears, therefore, to be entirely valid. This test is usually considered to be a powerful one, and one for which a sample size of eight is not unusual[4]. (Student[5] himself used sample sizes of as small as two.) Our original result, based on a two-tailed test, was $t=0.91$ with 7 d.f., therefore $P=0.39$. If, however, we do not pair the time intervals but do allow for the unequal variances in the two samples of time intervals and for the small sample sizes[6], we obtain $t=0.82$ with effectively 6 d.f., and so $P=0.45$. Alternatively, by simply testing the hypothesis that the average of the time intervals in the mass-extinctions series equals 33 Myr (which is the average time interval in the galactic series), we find $t=1.01$ with 7 d.f., so $P=0.35$. Clearly, even large differences in our test assumptions make little difference to the results.

Thus, we cannot reject the null hypothesis that the averages of the time intervals in the two series are equal. On the other hand, the tests cannot tell us how nearly equal the averages may be. A feeling for what can be safely rejected follows from a further consideration: with the same number of time intervals, another mass-extinctions time series might have a considerably different length. To demonstrate the consequences of adopting the original time series given by Raup and Sepkoski[3], in which the last nine mass-