of the ring complexes and dykes from air photographs forms a striking example of the value of photogeological interpretation. Work on the photographs continues at the headquarters of the Directorate of Colonial Geological Surveys.

K. V. STRINGER

Photogeological Section,
Directorate of Colonial Geological Surveys,
London.
April 17.

[1] Dixey, F., Campbell Smith, W., and Bisset, C. B., Nyasaland Geological Survey, Bull. No. 5 (1937).

Rapid Estimation of Standard Deviations

WHEN a series of numbers can legitimately be regarded as random samples from a normal distribution, the mean range within small sub-groups, multiplied by a factor depending only on group-size, gives a very good and easily calculated estimate of the standard deviation of the parent distribution, not much inferior in precision to that obtainable from the mean squared deviation[1]. A specially simple computation is obtained when the sub-group size is two.

Let the successive individual readings be A, B, C . . . X, Y, Z. The sum is required of all the ranges between adjacent numbers. This means that all the differences, $(A - B)$, $(B - C)$, and so on, are to be given a positive sign and added together. If B is greater than both A and C, it will have a positive sign twice. If B is less than both A and C, it will have a negative sign twice. But if B is bigger than one neighbour and smaller than the other, it will take one positive sign and one negative sign, and so will cancel out. The sum of the sub-group ranges, therefore, is twice the sum of the peak numbers (bigger than both neighbours) minus twice the sum of the trough numbers (smaller than both neighbours). This rule obviously breaks down for the terminal numbers A and Z. To deal with this, regard the series as an endless cycle, going back to A after reaching Z, and write the first two numbers again at the end, thus: X, Y, Z, A, B. It is then possible to see whether Z and A are peaks, troughs or intermediate, and to add, subtract or ignore them accordingly. When two or more adjacent numbers are identical, disregard all except the first and treat that by the regular rule. If there are N items in the series, there are also this number of sub-ranges. The factor for converting range into standard deviation for binary groups is 0·886, which can be taken as eight-ninths. Hence the following rule:

To estimate the standard deviation of a random series from a normal distribution, add all the peak numbers and subtract all trough numbers in the series with the first two members repeated at the end, divide by $\frac{1}{2}N$, and take off one-ninth.

The peak-minus-trough values can usually be found mentally, and may be written down and added. On a machine, the peak and trough totals can be found simultaneously, and the difference between them can then be obtained.

Range methods for standard deviation have been shown[1] to continue to give quite good approximations even when the distribution differs markedly from the normal. But the estimate may be seriously distorted if the data are not in random order. Any marked trend, whether upwards, downwards or cyclical, or any circumstance which causes adjacent values to be more closely alike or more unlike than would happen by chance, will produce bias in the estimate. In doubtful cases, there is a useful simple test of randomness. For large N, the number of peaks (or troughs) in a random series may be regarded[2] as normally distributed with mean $N/3$ and standard deviation $(N/22\frac{1}{2})^{1/2}$. The difference between actual and expected number of peaks is tested for significance in the usual way. If the discrepancy is D, there is no evidence, at the 95 per cent probability-level, of departure from randomness so long as D^2 is not larger than $N/5{\cdot}86$, or, to a sufficient approximation, not larger than $N/6$. When there are many 'ties' between adjacent observations, the number of these should be subtracted from N before the test for randomness is applied.

In practice, moderate departures from randomness do not invalidate the estimate of standard deviation, provided this method is recognized as giving only a quick approximation. It will often suffice to settle questions of significance without need for recourse to the maximally efficient but much more laborious computation based on sums of squares. It may also be used as an independent check of standard deviations calculated in the orthodox way. If this is done, the computer will soon gain experience of the effect of departure from randomness on the estimate of standard deviation. When the departure from randomness is serious, the method may still be applicable provided it is not difficult to shuffle the data so as to get nearer a chance order. It may suffice, for example, to read the data across rows instead of down columns. The method is specially useful for data recorded on control charts. It can also be used to get a pooled estimate from sets of observations presumed to have the same standard deviation though their means may differ, such as arise in biological assay. The peak–minus–trough figures are totalled over all the series, and treated as if they came from one series with ΣN observations.

To obtain the mean, make a separate total of the intermediate numbers (neither peak nor trough) and add to the peak and trough sums to get the grand total. All three totals can be computed simultaneously on the usual kinds of machine. An easier method, which gives the mean and a rather less precise estimate of the standard deviation, is to take the observations in non-overlapping pairs and to add separately the larger and the smaller numbers. Thus A and B would first be considered, then C and D, and so on, the smaller of each pair being accumulated on the left, and the larger on the right, of the machine. If N is even, the sum of the totals divided by N gives the mean, and the difference, divided by $\frac{1}{2}N$ with one-ninth deducted, gives the estimate of the standard deviation. If N is odd, the last observation may be omitted for computation of the standard deviation, or it may be paired with the first, the divisors being respectively $\frac{1}{2}(N - 1)$ or $\frac{1}{2}(N + 1)$. For computing the mean, every observation is, of course, taken once.

The methods described above are specially rapid on key-driven adding machines.

B. WOOLF

Institute of Animal Genetics,
West Mains Road,
Edinburgh 9.
March 17.

[1] Cf. Pearson, *Biometrika*, **37**, 88 (1950).
[2] Kermack and McKendrick, *Proc. Roy. Soc., Edin.*, **57**, 228 (1937).