THE REALITY OF REGULARITIES INDICATED IN SEQUENCES OF OBSERVATIONS

G. F. McEWEN

SCRIPPS INSTITUTION OF OCEANOGRAPHY, LA JOLLA, CALIFORNIA

SUPPOSE THAT A SEQUENCE of observations exhibits a striking regularity; for example, suppose that the values arrange themselves in their order of magnitude, either increasing or decreasing, or that they indicate a maximum or a minimum. Is the indicated regularity real, is it true of the sequence or statistical set sampled, or is it the result of sampling from a random sequence? That is if the process were repeated, is there reason to believe that approximately the same general result would recur? The more improbable it is that the indicated regularity could have arisen as a result of random sampling, the greater is the justification for regarding the indications of the sample as truly representative. Unless the probability of obtaining an indicated regularity is small, there is no evidence of the existence of an actual regularity of the indicated type.

First, assuming that only the sequence of individual numerical quantities is available, we may apply various tests based upon characteristics of a random sequence (see Besson, 1920; Clough, 1921, 1924; Crum, 1921; Woolard, 1925; Working, 1934; Kermack and McKendrick, 1936–37; as well as a series of later and more exhaustive contributions). For example, the number of maxima in a sequence of unrelated numbers is one-third of the number of terms. The departure of a sequence in any characteristic from that deduced for a random sequence suggests some systematic influence, and the evidence in favor of such an influence depends upon the extent of the departure and the number of terms. In general, such laws of chance sequences do not serve as criteria for proving the presence of some systematic influence, unless there are a large number of terms. Moreover, the results of attempting to determine the probability of obtaining a short sequence of terms having a striking appearance of regularity tend to be rather misleading.

Second, form a sequence of averages of groups of individual observations selected from a given sequence in a known systematic way. For example, a sequence might be repeated any number of times. Then averages of corresponding terms would form a composite sequence to be tested. The statistical significance of such a sequence of averages may be determined by comparing the variance of the individual observations computed directly with that derived from the variances of the averages. This analysis of variance principle can be extended to the general case in which values of the dependent variable are related to each of a number of correlated independent variables. In this problem of multiple curvilinear correlation a sequence of averages of the de-