LETTERS TO THE EDITOR

[The editors are not responsible for the views expressed in the letters.]

A NOTE ON NON-RANDOM FIELDS

In large scale sample surveys covering several hundreds of thousands of square miles the chief object is to make an estimate of a variate (which we may call u for convenience of reference) for the whole area. The usual procedure is to locate at random a large number of sample-units of a suitable size over the area, and measure the value of the variate u for each sample-unit. From this material it is possible to make an estimate of the mean value of u for the whole area under investigation.

The efficiency of the statistical design of the survey naturally depends on the way in which the variate u is distributed over the whole field or area under survey. From the point of view of sampling technique a field (in the sense of the area under survey) or any portion of a field is characterized by the type of distribution of the variate u over the field. This makes it necessary to consider the question of a systematic classification of different types of field. The object of the present note is to give a preliminary and brief account of work which is proceeding in this subject in the Statistical Laboratory.

It will be convenient for simplicity of discussion to consider each field to be made up of a large but finite number of space-units. There is no loss of generality in assuming these units to be of the same size. If each unit is considered to have a finite value of u including zero, then any such set of values of the variate will completely specify the distribution and hence the type of the field. We may also describe the field observationally with the help of contour lines separating regions differing by assigned levels of the variates. In drawing contour lines two alternative rules may be adopted, one based on the assumption of continuity, and the other admitting the possibility of discrete variation from point to point. This, however, is merely a question of detailed procedure, and does not affect the main argument.

The total number of separate regions (in the sense explained above) is thus an important characteristic of the field. It is clear, however, that any assigned number of regions will in general correspond to a number of different distributions. In fact each particular distribution may be treated as a micro-state and the observational configuration as a macro-state, so that a theoretical definition of the probability-weight of a macro-configuration can be set up on usual lines.

Considering all possible distributions corresponding to any macro-configuration we shall also have the usual expectation value of the number of separate regions which we may call the "random number of regions." Any field having a number of regions differing by less than a specified amount from the random number of regions may be defined as a "random field". Once the probability distribution of the number of regions can be determined, it will be then possible to construct a statistical scale showing the amount of deviation of any observed field from the random field ; that is, to construct a probability scale of non-randomness. It will be noticed that non-randomness as defined here may arise in two different ways, in the observed number of regions falling short of or exceeding the random number. This forms a convenient basis for classification.

In a field survey the sample-units may be of different sizes. It has been found that the variance between individual sample-units is, in general, a function of the size of sample-units. This variance function has a close connexion with the field type. For example, the variance function has the simple binomial or normal form, that is, varies inversely as the size of the sample-units in the case of random fields. This was experimentally verified in 1940 by extensive model sampling experiments details of which will be published in due course. It is likely that only random fields have a normal variance function, this is now being examined both theoretically and empirically. The connexion between the macro-configuration (and possibly its space derivatives in assigned directions) and the variance function is also being investigated in the case of non-random types.

Statistical Laboratory, P. C. MAHALANOBIS.
Calcutta, 21-5-1941.