

February 19th, 1936

Dear Monsieur Frechet,

I must begin my reply to your circular letter of February 8th by a purely verbal difficulty, namely as to the exact meaning which should be attached to the word "reperer", underlined in the middle of your second page. Should it be taken to mean simply to discover? In my ignorance I had recourse to my copy of Petit Larousse where I find "Marquer des repères", while a repere is "Marque faite a differentes pieces d'assemblage pour les reconnaitre et les ajuster plus facilement". This literal definition cannot be what is wanted, but I have missed the metaphor which should make it intelligible.

In the sense of simple discovery I have seen the correlation coefficient used with success, e.g. a soil physicist studying the plastic qualities of agricultural soils might find in the samples examined no association with the total percentage of calcium carbonate in the soil, but be delighted to find that a significant association appears when particles of calcium carbonate, too large to pass through holes 1mm in diameter, are excluded from the analysis. The search for such a significant association among the many variables, simple and compound, which may

be examined, is carried out not inconveniently in terms of the correlation coefficient, though certainly also not inconveniently in other ways. r is used, I think, in preference to other measures chiefly because the method of calculation is widely known and easily remembered, and because relatively precise tests of significance are easy to apply, owing to the availability of published tables.

The test of significance, i.e., the test of the hypothesis that there is no real association in the material from which the existing observations are regarded as a random sample, is necessarily equivalent to the test whether the coefficient b in the equation

$$y = a + bx$$

fitted to the data by least squares, as a means of predicting the value of y from that of x , differs significantly from zero. This is often the more appropriate and fruitful way of viewing the matter, but since the tests of significance are equivalent, and since, in the business of discovery, we are only concerned with significance and not necessarily with estimation, or even with knowing what we want to estimate, the correlation coefficient may be legitimately used in this way, even when its numerical value is without scientific meaning beyond that given it by its mathematical definition as a calculable function of observable quantities.

To say that this procedure is legitimate is, of course, not to say that it is the best that can be recommended. In particular cases other tests may be much more sensitive, that is to say they may be capable of demonstrating a significant association on the basis of fewer or less careful observations. This is an important advantage in statistical theory, but it is of little interest to the experimenter who, having found a significant correlation, is led, perhaps, to abandon a false theory by which he has hitherto guided himself, or to undertake a new line of research with different aims and different instruments.

I remain,

Yours sincerely,