

March 5th, 1936

My dear Professor Frechet,

You ask me to explain more clearly the kind of case I had in mind as an example of the detection or discovery of an association by means of the correlation coefficient. This is easy, for in both cases to which I referred I had the same criterion in mind, namely, whether the correlation coefficient does or does not exceed the least value which should be judged significant in regard to the number of observations available. Thus, if the soil physicist had 23 samples of different soils available, the properties of which he had examined and measured, he would probably choose the value .4227 as the least observed correlation which should be regarded as significant. The reason for this choice being simply that from uncorrelated and normally distributed populations this is the value that would be exceeded by chance just once in twenty trials. In order to judge a correlation to be insignificant, therefore, it is not necessary that the observed value

5/
2.12.63

should be very small, or near to zero. He would ignore it with confidence coming from so small a sample if it were between plus and minus .35. On the other hand, a value as high as .55 would only occur by chance once in several hundred trials in samples of this size in a population from which correlation was absent, so that the value of the correlation need not be near to unity in order to satisfy the experimenter that the value he has determined does demonstrate the ~~existence~~^{reality} of a real connection. It is the reality of this association which is vital to the experimenter, for on his confidence in its existence he is willing to spend, and perhaps to waste, the work he is capable of doing for perhaps several years.

What is important in this is that he should be provided with appropriate and sufficiently exact tests of significance. It is a matter of no consequence that in making these tests he should use the particular function of the observations (statistic) which we denote by r . In fact, we arrive at an exactly equivalent test if, ignoring the correlation coefficient, we calculate the regression of either one variate on the other by the familiar formulae of least squares and compare this with its standard error, using the exact procedure introduced by "Student"

for allowing for the limited number of degrees of freedom upon which the estimate of error has been based. It is, in fact, only the test of significance in its entirety which is of value to the experimenter, and not the particular statistic in which he happens to find it convenient to carry out the test.

I remain,
Yours sincerely,