

## "BIOMETRIKA."

By R. A. FISHER.

The last completed volume of BIOMETRIKA (Vol. x., April, 1914—May, 1915) contains a series of contributions of very exceptional interest to eugenists, presenting us, moreover, with the spectacle, rare in all the sciences, of an advance in theory, yielding, as soon as it is applied, practical and positive results of the first importance. In April appeared a short note, under the name of "Student," from which it appeared that that able statistician had, in consultation with Professor Pearson, devised a method by which a great mass of existing material may be made to give up its secrets. Hitherto it has been impossible to extract from such series of figures as annual price numbers, trade returns, and birth-rate statistics, the physical connections by which these quantities are related, owing to the effects produced on each or all of them by progressive change. There was no satisfactory way of eliminating the continuous and variable effect of time. "Student" shows that if this time factor may be represented by a power series  $bt + ct^2 + dt^3 + \dots + kt^m$ , so that

$$\begin{aligned}x_1 &= X_1 + bt_1 + ct_1^2 + \dots + kt_{1,m} \\x_2 &= X_2 + bt_2 + ct_2^2 + \dots + kt_{2,m}\end{aligned}$$

and so on for the  $n$  values of  $x$  corresponding, let us say, to  $n$  annual figures, then  $X_1, X_2, \dots, X_n$  will represent random variations in the quantities considered, with the time factor eliminated, and the way to find the correlation between a series  $X_1, X_2, \dots, X_n$  and another  $Y_1, Y_2, \dots, Y_n$ , is to find the correlation between the  $m^{\text{th}}$  differences of the two series,  $x$  and  $y$ .

The assumption of a power series for the expression of the time factor is, of course, open to question; but the method proposed by "Student" is to find in succession the correlation between the original figures, then that of the differences between successive figures, then of the second differences, and so on until a steady value for the correlation coefficient is reached. This is carried out in three examples. The correlation between Sauerbeck's index numbers for prices and the bankers' clearing house returns is for the crude figures  $-.33$ , but for the third, fourth and fifth difference the coefficient is small and apparently insignificant. The connection between marriage-rate and wages is of direct interest to the eugenist; here, starting with a large negative correlation  $-.52$ , we find from the second difference onwards values ranging closely about  $+.55$ , showing that the crude statistics as published had not only obscured the natural connection between these two things, but had reversed it.

In the November number two articles develop the thought of this short note. Dr. Anderson, of Petrograd, gives formulæ for the probable errors of the successive difference—correlations where there is no time-effect, and for the standard deviations of the successive differences where the differencing has been carried so far that the time-effect is eliminated. It is interesting that Dr. Anderson considers that the method is applicable when the time-effect cannot be represented by a power series. Miss Beatrice Cave and Professor Pearson exhibit an extensive trial of the new method. Taking Professor Montana's table of index values for ten different features of the economic development of Italy, a short series of 28 years, they tabulate the correlation coefficients of the crude figures and of the first six differences, taking each feature in turn with all the nine others, and with the "synthetic index," which is the arithmetic mean

of all ten. The result of this great labour, involving the calculation of seven correlation coefficients for each of the fifty-five pairs of variables, is a striking success in showing the importance of the new process. The crude figures give correlations, of which the lowest is the very high value  $+ .885$ ; as the difference correlations are brought out remarkable changes take place. The genuine correlations are separated from those which are spurious. The correlation between the indices for the gross consumption of tobacco and for savings banks' returns is  $.984$ , from which one might imagine that private expenditure was so well regulated that no one would spend the smallest sum on tobacco without laying by a fixed proportional amount in savings. But for the third and succeeding differences we read  $-.327$ ,  $-.380$ ,  $-.402$ ,  $-.431$ , showing that, as we should have expected, when expenditure on tobacco is great, savings are correspondingly diminished. It would be interesting to know how the regressions worked out on a uniform currency basis. The Variate Difference Correlation Method has evidently a great future. We do not yet know how much precious information may be hidden, masked by the time-effect, in existing vital statistics. On the economic side we may note the significant remark "The very superficial statements, so frequently met with, that such and such variates, both changing rapidly with the time, are essentially causative, will doubtless cease to have scientific currency directly the method of variate differences is fully appreciated. We shall no longer assert that the fall in the phthisis death-rate can be off-hand causatively associated with the contemporaneous rise in the number of persons dying in institutions, or that the increased expenditure on business is necessarily a measure of increased national prosperity" (p. 353).

The first fruits of the application of the new method to vital statistics appear in the May number as "Further Evidence of Natural Selection in Man," by Miss Ethel Elderton and Professor Karl Pearson. The correlations are found between death-rates of the same group of infants in successive years of their life. Infantile death-rate is affected by the time factor, being continuously influenced by industrial and administrative changes, so that the figures drawn from the Registrar-General's reports give positive correlations for the death-rates of successive years. Neglecting the effects of changing conditions, this would lead to the surprising conclusion, eagerly drawn by those whose sentiment outruns their reason, that the agencies of infantile death exert a weakening effect upon the survivors, and not as one would expect, a raising of the average level by the removal of the weakest. Pearson and Snow had both shown conclusively that the average resisting power is as a matter of fact raised, and that a high death-rate in the early years of childhood is the natural effect of a low death-rate in infancy. But their methods of removing the time factor were not so complete, and their results correspondingly less conclusive than in the present paper.

The death-rates were worked out by comparing the number of children born in any year with the number of deaths under one year of age in that year, with the deaths between one and two in the succeeding year, and so on. It is possible that the results are somewhat affected not only by immigration and emigration, but by the grouping into years necessitating the consideration of the death-rate of not exactly the same group of babies. The variate difference method would seem to multiply any error due to babies getting into the wrong year, but it should be noticed that these effects would tend merely to blur the results and decrease the correlations obtained; these are so uniform and definite as to remove any doubt as to their significance. The main results are as follows. The correlation between the death-rates in successive years of life up to five is very approximately  $-.7$  for both boys and girls. This is a remarkably high value, which is approached regularly by all eight series of numbers. One can only conclude that natural selection is actively at work in civilised

countries. It is important, in view of the limitation of families, that there should be no doubt on this point. The partial correlations for a two-year interval, when the intervening year is rendered constant were also found. For the first and third year this is about  $-.48$ , for the other two pairs about  $-.23$  and  $-.25$ . So that the effects of selection are well marked even in this case.

We may confidently hope that many other controversial matters will be made clear by the new means at our disposal. There must be many reactions between vital and economic data which take some years to take effect. Much may be done by finding such lagging correlations. Again, there would seem to be a prospect of the establishment of the fundamental differential equations of sociology by correlating differences of different orders.