Vol. I. Part I. October 1901

BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS

EDITED
IN CONSULTATION WITH FRANCIS GALTON
BY
W. F. R. WELDON KARL PEARSON
AND
C. B. DAVENPORT

CAMBRIDGE
AT THE UNIVERSITY PRESS
LONDON: C. J. CLAY AND SONS, AVE MARIA LANE
AND H. K. LEWIS, GOWER STREET
NEW YORK: THE MACMILLAN COMPANY
LEIPSIC: BROCKHAUS

Price Ten Shillings net.

IGNORAMUS, IN HOC SIGNO LABOREMUS.

Charles Darwin

BIOMETRIKA.

EDITORIAL.

(I.) *The Scope of Biometrika.*

IT is intended that *Biometrika* shall serve as a means not only of collecting under one title biological data of a kind not systematically collected or published in any other periodical, but also of spreading a knowledge of such statistical theory as may be requisite for their scientific treatment.

A very few years ago, all those problems which depend for their solution on a study of the differences between individual members of a race or species, were neglected by most biologists. The complexity of organic structure is so great, and the number of distinguishable forms so enormous, that morphologists were obliged to simplify their conceptions by constructing for every species an ideal type, to which the individuals composing it conform with more or less exactness, and to neglect those deviations from the type which actually occur. Such simplification was not only justifiable, but absolutely necessary for many purposes; it has rendered enormous service to biology in the past, it does so still, and will continue to do so; nevertheless, there are many problems which cannot be dealt with by its aid.

The starting point of Darwin's theory of evolution is precisely the existence of those differences between individual members of a race or species which morphologists for the most part rightly neglect. The first condition necessary, in order that any process of Natural Selection may begin among a race, or species, is the existence of differences among its members; and the first step in an enquiry into the possible effect of a selective process upon any character of a race must be an estimate of the frequency with which individuals, exhibiting any given degree of abnormality with respect to that character, occur. The unit, with which such an

enquiry must deal, is not an individual but a race, or a statistically representative sample of a race; and the result must take the form of a numerical statement, showing the relative frequency with which the various kinds of individuals composing the race occur.

As it is with the fundamental phenomenon of variation, so it is with heredity and with selection. The statement that certain characters are selectively eliminated from a race can be demonstrated only by showing statistically that the individuals which exhibit that character die earlier, or produce fewer offspring, than their fellows; while the phenomena of inheritance are only by slow degrees being rendered capable of expression in an intelligible form as numerical statements of the relation between parent and offspring, based upon statistical examination of large series of cases, are gradually accumulated.

These, and many other problems, involve the collection of statistical data on a large scale. That such data may be rendered intelligible to the mind, it is necessary to find some way of expressing them by a formula, the meaning of which can be readily understood, while its simplicity makes it easy to remember. The recent development of statistical theory, dealing with biological data on the lines suggested by Mr Francis Galton, has rendered it possible to deal with statistical data of very various kinds in a simple and intelligible way, and the results already achieved permit the hope that simple formulæ, capable of still wider application, may soon be found.

The number of biologists interested in these questions, and willing to undertake laborious statistical enquiries, is already considerable, and is increasing. It seems, therefore, that a useful purpose will be served by a journal especially devoted to the publication of statistical data, and of papers dealing with statistical theory. Many persons are deterred from the collection of biometric data, by the difficulty of finding such a means of publishing their results as this journal will afford, and those results which are published frequently lose much of their value because the data on which they are based are withheld, or because they are isolated in publications largely devoted to other forms of investigation. Further, *Biometrika* will endeavour to introduce a uniformity of statistical treatment, terminology, and notation, so that results obtained by different investigators on different types of life may be easily and effectively compared.

Biometrika will include (*a*) memoirs on variation, inheritance, and selection in Animals and Plants, based upon the examination of statistically large numbers of specimens (this will of course include statistical investigations in anthropometry); (*b*) those developments of statistical theory which are applicable to biological problems; (*c*) numerical tables and graphical solutions tending to reduce the labour of statistical arithmetic; (*d*) abstracts of memoirs, dealing with these subjects, which are published elsewhere; and (*e*) notes on current biometric work and unsolved problems. It is proposed to include memoirs written in English, German, French, or Italian.

(II.) *The Spirit of Biometrika.*

It is almost impossible to study any type of life without being impressed by the small importance of the individual. In most cases the number of individuals is enormous, they are spread over wide areas, and have existed through long periods. Evolution must depend upon substantial changes in considerable numbers and its theory therefore belongs to that class of phenomena which statisticians have grown accustomed to refer to as *mass-phenomena.* A single individual may have a variation which fits it to survive, but unless that variation appears in many individuals, or unless that individual increases and multiplies without loss of the useful variation up to comparatively great numbers—shortly, until the fit type of life becomes a mass-phenomenon, it cannot be an effective factor in evolution. The moment this point is grasped, then whether we hold variation to be continuous or discontinuous in magnitude, to be slow or sudden in time, we recognise that the problem of evolution is a problem in statistics, in the vital statistics of populations. Whatever views we hold on selection, inheritance, or fertility, we must ultimately turn to the mathematics of large numbers, to the theory of mass-phenomena, to interpret safely our observations. As we cannot follow the growth of nations without statistics of birth, death, duration of life, marriage and fertility, so it is impossible to follow the changes of any type of life without its vital statistics. The evolutionist has to become in the widest sense of the words a registrar-general for all forms of life. When he cannot observe and measure in Nature, then he must experiment on "populations" within the laboratory. But few biological laboratories have the space or the resources needed for dealing with the vital changes of populations, still less do the means at the disposal of individuals suffice for carrying out extensive experiments of this character. Much has been done and undoubtedly more will be done by the Marine Biological Laboratories for the study of mass-phenomena, but what is urgently needed is the establishment of a well-equipped Biometric Farm Laboratory, where breeding and survival experiments on large numbers could be carried out with ample room and care and, when necessary, for long periods. To this point we hope to return, and we shall not cease to urge its importance*.

But if we have thus to deal with a mass-phenomenon, may we not ask how it came about that the founder of our modern theory of descent made so little appeal to statistics? An illustration may aid us; the structure of our present theory of the moon is the creation of Newton, using his characteristic geometrical methods. But the practical astronomer in all lunar investigations to-day applies the analytical methods subsequently devised by the French mathematicians. The characteristic bent of Charles Darwin's mind led him to establish the theory of descent

* The failure of an attempt in this direction made a few years ago was, we believe, largely due to the fact that some of its supporters had not realised that the problem of evolution is a problem in the vital statistics of large numbers.

without mathematical conceptions; even so Faraday's mind worked in the case of electro-magnetism. But as every idea of Faraday allows of mathematical definition, and demands mathematical analysis in its modern statement, so every idea of Darwin—variation, natural selection, sexual selection, inheritance, prepotency, reversion—seems at once to fit itself to mathematical definition and to demand statistical analysis. Nor was the statistical conception itself entirely wanting in Darwin's work. *The Cross and Self-Fertilisation of Plants* forms a splendid collection of statistical observations and experiments which offers many points of departure for further statistical research*. That Darwin's mind did not work easily in mathematical lines is, perhaps, best evidenced in the passage of a letter of 1857 to Sir John Lubbock written when Darwin was dealing with the statistics of varieties in species as deduced from *Floras*:

> You have done me the greatest possible service in helping me to clarify my brains. If I am as muzzy on all subjects as I am on proportion and chance,—what a book I shall produce! (*Life*, II. p. 104).

But that he realised the importance of the statistical method for his investigations is evidenced not only by this very passage, but by several others. Thus considering the *variation* of our common species he writes in 1846:

> Andrew Smith once declared he would get some hundreds of specimens of larks and sparrows from all parts of Great Britain, and see whether, with finest measurements, he could detect any proportional variations in beaks or limbs, etc. This point interests me from having lately been skimming over the absurdly opposite conclusions of Gloger and Brehm (*Life*, II. p. 35).

Andrew Smith indeed missed the opportunity of being a veritable biometric pioneer!

Elsewhere Darwin recognises the importance of determining the variability of skeletons by measuring limbs (*Life*, II. p. 50). But, perhaps, the strongest evidence of his consciousness that biometry offers the only possible solution of problems of *inheritance* occurs in the words:

> I write now to say that I have been looking at some of our mongrel chickens, and I should say *one week old* would do very well. The chief point which I am, and have been for years, very curious about, is to ascertain whether the *young* of our domestic breeds differ as much from each other as do their parents, and I have no faith in anything short of actual measurement and the Rule of Three (*Life*, II. p. 51).

These words prove fully Darwin's consciousness not only of the need of measurements, but also of arithmetical work upon such data in the case of heredity. They may well serve as a motto for *Biometrika* and for all biometricians: *I have no faith in anything short of actual measurement and the Rule of Three.*

It is not a mere formal clothing of biological conceptions with mathematical symbols that is here indicated, or that we are considering, when we say that all

* See for example Mr G. U. Yule's use of Darwin's data in his recent memoir "On the Association of Attributes in Statistics," *Phil. Trans.* Vol. 194, A., 258.

Darwin's ideas fit themselves to algebraic definition. On the contrary—exactly as in the like case of the mathematical treatment of Faraday's conceptions of electromagnetism—the symbolic analysis widens our notions, it leads us at once to new points of view and it directly suggests—perhaps this is its most important advantage—fresh points for observation and novel directions for experimental research.

The danger will no doubt arise in this new branch of science that—exactly as in some branches of physics—mathematics may tend to diverge too widely from Nature. The biologist, the mathematician and the statistician have hitherto had widely differentiated fields of work. Each one of these fields is full of pitfalls, and when the worker amid living types wanders among symbolic forms, the mathematician by profession must give him a helping hand if he stumbles over a determinant or gets entangled in a differential. A like patience must be extended by the biologist to the mathematician when he makes blunders at which the morphological tyro would smile. As Mr Francis Galton said a few years ago, for these new problems we want a scientific firm with a biologist and a mathematician as acting partners and a logician as a consulting partner. Patient endeavour to understand each other's methods, and to bring them into harmony for united ends and common profit—this is the only method by which we can win for biometry a recognised place in the world of science and in the accepted academic curricula of the universities. The day will come—is, perhaps, already dawning with the younger workers—when we shall find mathematicians who are competent biologists, and biologists who are competent mathematicians, but our universities both as to teachers and laboratories are not yet adapted for the training of such men, and for some time to come we can in the main only hope for effective partnership and not for the all-round biometrician. We have a splendid, almost untraversed field to work in, and a great task to perform in winning not only full recognition from the scientific world, but public support for our work. If these conditions are kept in view the diverse degrees of mathematical and biological knowledge exhibited in our pages will not oppress our readers. We shall publish careful biometric observations, even if they be accompanied by only the most elementary statistical treatment; we shall look forward to our mathematical workers supplementing such fundamental observations by more elaborate statistical calculations. For this reason we shall not only print as copious observational and experimental data as possible, but endeavour to form a manuscript collection of such data available for further research. We hope that every number of *Biometrika* will present statistical material ready for the mathematician to calculate and to reason upon. All such investigations ancillary to data appearing in our pages we shall receive gladly and publish at the earliest opportunity.

On the other hand the biologist will find in our pages algebraic analysis which may repel him. We would still ask his attention for the general conclusions and for the formulæ reached by the mathematician. The biologist will find that they frequently suggest observations and experiments which he alone is in a position to undertake satisfactorily. We shall aid the more arithmetical part of his work by

diagrams and numerical tables wherever it seems possible. In this manner we hope that *Biometrika* will provide material for both branches of science; that it will not only publish valuable biometric and statistical researches, but serve as a store-house of unsolved problems for both unemployed biologist and mathematician. We trust that bringing these men together may widen the activity of both.

Many of the problems of biometry can only be approached from the standpoint of the cooperative collection of material and reduction of statistics, and in this respect we shall strive to form a link between scattered biometric workers. *Biometrika* will ask for aid in cooperative work, and be at all times ready to publish requests for aid, and the forms and schedules which our contributors desire to be circulated or filled in. In this way it may be hoped that a guild of qualified collectors and workers may be gradually formed to whom appeal may be made for collecting, counting and observing. There are many men and women with the necessary training, scattered about this and other countries, who without having the opportunity for initiating original work are not only competent but glad to assist with collecting-box, camera or pencil. From such workers the Editors will be glad to hear and will endeavour to put them in touch with those desiring their aid.

Extensions, corrections, criticisms of the results published in our pages we shall heartily welcome whatever be their source. We expect to receive stalwart blows as well as to give them. All we shall demand in this respect is the chivalry which is needful in scientific controversy, which while combating error does not discourage honest endeavour. The most fertile men of science have made blunders, and their consciousness of such slips has been retribution enough; it is only their more sterile critics who delight to dwell too often and too long on such mistakes. In science, both in symbolic analysis and in our knowledge of Nature, we are very ignorant; we do not pretend that biometry will revolutionise our ideas of life. All we claim is that in certain aspects of biological research, biometry is an instrument which can aid us effectively in our gropings after truth. Only let the spirit in which it is used be that of the master-mind, the ideal so well and faithfully portrayed in the form and features on our frontispiece. *Ignoramus; in hoc signo laboremus!*

BIOMETRY.

By FRANCIS GALTON.

This Journal is especially intended for those who are interested in the application to biology of the modern methods of statistics. Those methods deal comprehensively with entire species, and with entire groups of influences, just as if they were single entities, and express the relations between them in an equally compendious manner. They commence by marshalling the values in order of magnitude from the smallest up to the largest, thereby converting a mob into an orderly array, which like a regiment thenceforth becomes a tactical unit. Those to whom these considerations are new, will grasp the results more easily by thinking of the array in its simplest, though not necessarily in its most convenient, form for mathematical treatment. Let them conceive each value to be represented by an extremely slender rod of proportionate length, and the rods to be erected side by side, touching one another, upon a horizontal base. The array of closely packed rods will then form a plane area, bounded by straight lines at its sides and along its base, but by a flowing curve above, which takes note of *every one* of the values on which it is founded, however immense their multitude may be. The shape of the curve is characteristic of the particular group of values to which it refers, but all arrays have a family resemblance due to similarity of origin; they all drop steeply at one end, rise steeply at the other, and have a sloping back. An array that has been drilled into some such formation as this, is the tactical unit of the new statistics. Its outline is expressed by a general formula whose constants are adapted to each particular case, and, being thus brought within the grip of mathematics, the internal relations of an array and their relations to those of any other array can be expressed in exact numerical forms. The new methods occupy an altogether higher plane than that in which ordinary statistics and simple averages move and have their being. Unfortunately the ideas of which they treat, and still more the many technical phrases employed in them, are as yet unfamiliar. The arithmetic they require is laborious, and the mathematical investigations on which the arithmetic rests are difficult reading even for experts; moreover they are voluminous in amount and still growing in bulk. Consequently this new departure in science makes its appearance under conditions

that are unfavourable to its speedy recognition, and those who labour in it must abide for some time in patience before they can receive much sympathy from the outside world. It is astonishing to witness how long a time may elapse before new ideas are correctly established in the popular mind, however simple they may be in themselves. The slowness with which Darwin's fundamental idea of natural selection became assimilated by scientists generally, is a striking example of the density of human wits. Now that it is grown to be a familiar phrase, it seems impossible that difficulty should ever have been felt in taking in its meaning. But it was far otherwise, for misunderstandings and misrepresentations among writers of all classes abounded during many years, and even at the present day occasional survivals of the early stage of non-comprehension make an unexpected appearance. It is therefore important that the workers in this new field who are scattered widely through many countries, should close their ranks for the sake of mutual encouragement and support. They want an up-to-date knowledge of what has been done, and is doing, in it. They seek for opportunities of receiving judicious help from one another, sometimes in circulating questions, sometimes in discussing the preliminaries of new plans of campaign. Immense labour has too often been wasted in statistical research through a mistaken judgement of the value and real significance of the data employed. The fresh opinion of skilled onlookers is the safest test of the value of materials and affords a ready means of obtaining timely warning of the presence of vitiating conditions before an inquirer commits himself to any new statistical enterprise. Every investigator stands in need of expert criticism, for no pursuit runs between so many pitfalls and unseen traps as that of statistics.

This Journal, it is hoped, will justify its existence by supplying these requirements either directly or indirectly. I hope moreover that some means may be found, through its efforts, of forming a manuscript library of original data. Experience has shown the advantage of occasionally rediscussing statistical conclusions, by starting from the same documents as their author. I have begun to think that no one ought to publish biometric results, without lodging a well arranged and well bound manuscript copy of all his data, in some place where it should be accessible, under reasonable restrictions, to those who desire to verify his work. But this by the way. There remains another cogent reason of a very practical kind for the establishment of this Journal, namely that no periodical exists in which space could be allowed for the many biometric memoirs that call for publication. Biometry has indeed many points in common with Mathematics, Anthropology, Zoology, Botany, and Economic Statistics, but it falls only partially into each of these. An editor of any special journal may well shrink from the idea of displacing matter which he knows would interest his readers, in order to make room for communications that could only interest or even be understood by a very few of them. I am tempted to illustrate, or rather to over-illustrate, the coldness of welcome often afforded to a new departure in science, by an anecdote concerning the cause that really led to the foundation of the Geological Society

of London. I have rarely related it in conversation, fearing to give pain to some one, and I have never done so in print; neither can I find that any version of it has been published by others. But now that nearly a century has slipped past since the event, there can be no harm in digging up and bringing to light a buried but amusing historical fact.

The story was told me long, long ago, in the 'forties, by Mr George Bellas Greenough, F.R.S. I was then an eager youth fresh from college, and he an elderly man; it was as follows. In 1806–7, when Geology was in its infancy and travellers were scarce owing to European wars, Mr Greenough and a few young friends compiled a list of questions with the view of ascertaining how far the facts of Nature might agree with the competing geological theories current in those days. Sir Joseph Banks was the President of the Royal Society at that time, an office which he exercised despotically for 43 years (1777–1820), becoming almost an autocrat over English scientific men. So it was to him that Mr Greenough and his young friends naturally went. They brought their questions and begged that copies of them might be circulated under official sanction among suitable persons, including foreign correspondents of the Royal Society. Sir Joseph was sometimes gracious in mood, frequently the reverse, and on this occasion he might be described as bearish. Not content with an emphatic "no," he dismissed them with words to the effect (in almost those very words, if my memory does not deceive me) that a few fools could ask more questions in half an hour than wise men might answer in years. The deputation departed, ready to burst with suppressed fury, and the moment they were quit of the house, agreed to circulate the questions on their own responsibility, which considering the persons and circumstances was an act of rare audacity. Out of this impromptu coalition, aided by a multitude of elsewhere recorded circumstances, the Geological Society was evolved, with Mr Greenough as its first President. (The official account of its origin is judiciously reticent, but not inconsistent with this little piece of history. It will be found in the preface to the first volume of its *Transactions*, published in 1811.) It is not in the least my intention to insinuate that Biometry might be served by any modern authority in so rough a fashion, but I offer the anecdote as forcible evidence that a new science cannot depend on a welcome from the followers of older ones, and to confirm the former conclusion that it is advisable to establish a special Journal for Biometry.

The primary object of Biometry is to afford material that shall be exact enough for the discovery of incipient changes in evolution which are too small to be otherwise apparent. The distribution of any given attribute, within any given species, at any given time, has to be determined, together with its relations to external influences. This affords a standard whence departures may be measured and the direction and rate of their progress ascertained. Evolutionary changes are exceedingly slow as a rule, but supposing that a thousand years or thirty generations of mankind, would suffice in some particular case for some conspicuous alteration in a species, exact measurements ought to discover its progress well within the limits

of a human lifetime. Moreover the forms by which distribution is expressed in the new method are excellently fitted to bring to light any survivals of a less advanced type, which may serve as evidence of recent change. Also they quickly indicate incipient changes, through their power of isolating aberrant forms, and then of measuring the degree in which any of these may be favoured by natural selection. The organic world as a whole is a perpetual flux of changing types. It is the business of Biometry to catch partial and momentary glimpses of it, whether in a living or in a fossil condition, and to record what it sees in an enduring manner. It is an after-process to combine those glimpses into a continuously changing scene, much as some tumultuous procession is made to live and move again by means of a "biograph." Each biometric investigation may be compared to a solitary boring in a level plain, whose underlying geology has to be ascertained. A comparison of the cores brought up, will supply evidence of the depths of each of the buried strata and will justify many interpolations of unseen portions between the borings. For instance, it may not require many investigations to establish statistical laws of heredity on a secure basis, by ascertaining the limits within which those that have been already observed may hold good in a moderate number of widely different types of plant and animal life. Biology could soon be raised to the status of a more exact science than it can as yet claim to be, if each of many biometricians would thoroughly work out his own particular plot, although those plots may be very far indeed from occupying the whole of the area that admits of being directly explored.