

# SANKHYĀ

## THE INDIAN JOURNAL OF STATISTICS

*Edited by* : P. C. MAHALANOBIS

---

VOL. 20, PARTS 3 & 4

DECEMBER

1958

---

### THE SCOPE OF BIOLOGICAL STATISTICS<sup>1</sup>

*By* J. B. S. HALDANE

*Indian Statistical Institute, Calcutta*

Statistics are always based on counting, and often on measurement. One may merely count all the individuals of a certain type, such as rice plants, in a certain area. One may count individuals of several well marked types, such as members of different castes or blood groups, fish of different species in a tank, bean pods containing zero to eight beans, and so on. Or one may measure, and then group the individuals, for example fish weighing between 1 and 1.1 kgm, or households with annual incomes between 1500 and 2000 rupees. But the fundamental data are the numbers belonging to different classes.

The individuals counted may be objects or events. Thus you may want to know how many buses there are in Calcutta, or how many passed a particular point in a given direction in every 5 minutes during 24 hours. If you do this you will count the same bus several times in the day.

A biologist may be interested in objects and events with very different scales in time and space. He may want to estimate the numbers of various kinds of cells in a cubic millimetre of human blood. This contains about 5 million red corpuscles and 10,000 white corpuscles which can be classified into about 6 different types. Or he may be concerned with the human population of India or of the world. Again he may be counting nervous impulses of which 50 or so can be detected in a nerve fibre per second, or finding from the fossil record the mean duration a genus (about 8 million years for carnivora, 40 million for mollusca). His methods will, of course,

---

<sup>1</sup> This is Professor J. B. S. Haldane's inaugural lecture as Research Professor in the Indian Statistical Institute. It was delivered at a public function held in the Institute on 4 October 1957 at which Professor S. N. Bose, Vice-Chancellor of Visva-Bharati presided.—*Editor*.

be quite different in different cases. For the small objects he will need a microscope, for the rapid events a record on a rapidly moving sensitive film. And so on. This means that a statistician must know something about methods of measurement. He must know that a beam of electrons acts as a lever without inertia in recording rapid processes, and radioactivity as a very slow clock in recording slow ones. He must know that the limits of accuracy in measuring small objects are given by wave lengths, those of visible light in microscopy, of electrons considered as wave packets in electron microscopy. If he does not understand the methods of measurement, he may grossly exaggerate the variation of what is measured. When he has data on a sample he will look for representative parameters such as means or medians, measures of spread round them, and measures of relation between different qualities or measures such as correlations.

Let us take some examples at the microscopical level. The number of chromosomes observable in the dividing nuclei of a plant or animal is usually constant in the germ line, that is to say the cells which give rise to future generations. It is halved when the gametes, for example spermatozoa or pollen grains, are formed. The large majority of nuclear divisions forming them are regular in normal organisms. However, in a fraction of the order of one per thousand the process may go wrong, a gamete being produced with a chromosome too few or too many. But in hybrids between species these errors are much commoner, and one can sometimes classify gametes according to the number of chromosomes which they contain. As however the chromosomes may be as little as  $1\mu$  (.001 mm) in length, errors are unavoidable, and certain workers tend only to count nuclei giving a clear picture. This obviously leads to bias. Again the diameter of a human red blood corpuscle (about  $7.5\mu$ ) is somewhat variable. The variance may be greatly increased in disease. But diameters are usually estimated from photographs, and we still know too little about errors of measurement. For example, Attfield (1951) in my laboratory found a mean corpuscular diameter of  $7.65\mu$  in normal, and  $8.59\mu$  in congenitally anaemic mice. But the coefficients of variation were 9.2% in the normals and 13.3% in the anaemics. The mean was increased by only 12%, the coefficient of variation by 45% or the variance by 165%. As we do not know how much of the variation is observational, the true increase is certainly greater.

The objects counted are usually individuals, but sometimes classes, sometimes both. An example of classes which are counted is given by species. For example about 200,000 species of flowering plants have been described. 21,000 of these are found in India. The only state with a larger number is Brazil. The U.K. has only about 1500, many known to be recent importations. Of course the number of species is nothing definite. But it is easier to count species than banana or banyan trees. Again if you catch  $N$  insects with a light trap they will belong to  $n$  species. The expected value of  $n$  is not far from  $k \log N$ , where  $k$  is a measure of the diversity of the insect population sampled. Comparative studies on these lines are needed. Little has been done in tropical countries.

The classical biometry of Karl Pearson was largely concerned with measurements such as stature, head length, strength of grip, and so on, though it must be said that Pearson's work on fertility was based on the counting of offspring. This biometrical work is certainly worthy of continuation. It is particularly important economically, since economically important biological characters, such as the milk yield of a cow, or the rice yield in grams per plant or per hectare are usually continuous variables. Before we continue I wish to consider variance as a biologist rather than a statistician. If a character has a very small coefficient of variation in a population we may expect that it is important. Deviants may be weeded out by human agency. They may be weeded out by natural selection, as for example Rendel (1943) found that very light or very heavy duck's eggs had less probability of hatching than those of median weight. Or it may be regulated physiologically as is the human temperature. An increased variance may mean a breakdown of natural or artificial selection or of physiological regulation. For example Callan and Spurway (1951) studied spermatogenesis in hybrid newts. They found that the mean number of chiasmata (junctions between homologous chromosomes) was considerably diminished. But the variance was greatly increased. That is to say the mechanism of formation of the nuclei of spermatozoa was no longer being controlled at all closely. A decrease in a correlation coefficient usually implies increased variance in an index such as a ratio of two lengths, which can often be used as a measure of shape. Thus if the length and breadth of leaves are more highly correlated on one tree than on another, this means, roughly, that their shape is more constant, or more precisely regulated.

Pearson investigated homotypy, that is to say he discovered what fraction of the variance of a measurable or countable character, such as the number of ridges on a seed capsule, a shape parameter of a leaf, or the length of an egg, in a population, was due to variation within so called individuals. (I emphasize the unsatisfactory nature of the word individual. Palm-trees of most species seem to be individuals. They cannot be propagated by division. But most trees would more accurately be described as individuals). This work requires extension, because, as S. K. Roy has shown, the mean of a character in different flowers of the same plant may vary during a season, as may its variance. In fact we shall have to think of each "individual" plant as defining a population of leaves, flowers, fruits, and so on, and a group of plants as a population of these populations. It might be possible to calculate generalized distances between pairs of plants, and to find the distribution of these in a large population of plants.

We may perhaps expect quicker returns from mere counting, for example counts of the numbers of seeds on different plants; and I hope that someone may do for Indian plants what Salisbury (1942) has done for British plants in this respect. One can also study natural selection, that is to say the survival and fertility of organisms differing in qualitative or measurable characters. An extremely important quantitative character is resistance to insecticides and other poisons. Workers have usually contented themselves with estimating the median lethal dose of a poison, and perhaps, though much more roughly, the dose killing 99%. It is no doubt satisfactory, for

the moment, to reduce the population of mosquitoes in an area to 1% of its former value. But this involves intense selection in favour of resistance to the insecticide used, and experience has shown that in many cases the Darwinian view that such selection will increase the resistance in future generations turns out to be correct.

The classical biometricians were interested in heredity, and therefore in characters which they believed were little affected by the environment. Geneticists have naturally shown the same bias. Human blood group membership is an ideal character for geneticists, because no case is recorded with certainty in which it has changed during a human lifetime. If, however, one is interested in improving the environment one should clearly look for characters which are mainly affected by it. No measurable human characters are entirely determined by the environment, as is, for example, the language spoken at the age of five years. However, it is clear that the weight of adults, or better the weight divided by the cube of the height, is more responsive to nutritional differences than is the height. Recent American workers have chosen such measures as the thickness of a skin fold as measures of nutritional status. The urinary excretion of vitamins is another. Work on such characters will be particularly valuable in India.

Any biometrical work which does not allow of the comparison of relatives has a limited value. Though once the inheritance of such a character as blood group membership is well established, one need not make such a comparison in each population studied. The methods of non-human genetics, however, differ fundamentally from those of classical biometry in being experimental. If a geneticist suspects that a rare character is recessive he immediately mates two animals or plants bearing it, or self-fertilizes a plant. Only if all the progeny resemble their parent or parents can he conclude that it is recessive. This cannot be done with a supposedly recessive human character such as microcephalic idiocy. Animal and plant genetics almost always involve an alteration in the previous mating system of the population studied.

In consequence they involve novel statistical methods. Let me take an example. A sample of a human population includes  $a$  members of blood type  $M$ ,  $b$  of type  $MN$ , and  $c$  of type  $N$ . On the null hypothesis that all types have equal survival and fertility, and that mating is at random as regards these characters, we can easily show that the expectation of the statistic  $4ac - b^2 + b$  is zero. But it is not quite a trivial problem to calculate its exact sampling distribution, and to establish confidence or fiducial limits. This problem is however trivial compared with that which faces an experimental geneticist. He finds  $D$  dominants and  $R$  recessives in a sample, having previously shown to his own satisfaction that the characters in question are dominant and recessive. He mates  $n$  dominants to recessives and obtains progeny. The number of progeny in the  $i$ -th of  $a$  families containing no recessives is  $d_i$ , the number in the  $i$ -th of  $b$  families containing at least one recessive is  $d'_i$  dominants and  $r'_i$  recessives. From these figures it is required to construct statistics which will enable us to test various null hypotheses, including the hypothesis that dominants and recessives are equally viable, and that mating was at random in

## THE SCOPE OF BIOLOGICAL STATISTICS

the population sampled. I have not solved this problem, though I have given an approximate solution. I do not think that the same solution would satisfy Fisher and Neyman.

Besides changing the mating system we may try to induce mutations with radioactive substances or otherwise, and to measure the frequency with which we do so. The method which I have designed for this purpose (Haldane 1958) and which is now being used by Dr. Carter at Harwell, will, I think, measure this frequency in mice within 10% or so. Hence until a hundred induced mutations have been recorded the inaccuracy of various assumptions which I have made will not matter much. But if greater accuracy is desired, a number of extremely difficult problems will have to be solved.

A very important field of biological statistics centres round life tables. These are often based on such large numbers that sampling errors are negligible compared with other errors which beset all official statistics, and are not wholly avoided even in sample surveys. On the other hand a biologist may obtain a life table based on only a hundred animals, while being sure, either that there were no errors, or that three of the hundred were accidentally killed on known dates. The sampling theory appropriate to such data only exists in part.

Still less satisfactory are the methods for dealing with events whose probability varies with the animal's age. Within a pure line of mice, that is to say, a genetically homogeneous group of mice of the same sex, kept under fairly uniform conditions, cancer may be negligible till after the median age of death. The force of mortality from cancer, that is to say the probability that a given individual will die of cancer in a short period, then rises, and may apparently reach a more or less constant value which may be of the order of 1 to 2 percent per day. One can calculate, for a given stock under given conditions, the median age of death from cancer, that is to say the age at which half the mice would be dead if all other causes of death were eliminated. But in doing so one must make the assumption that causes of death are independent, in the sense that those mice which died, say of kidney disease, were no more and no less likely than others to die of cancer within a given time if we could have prevented the kidney disease. This is rather unlikely in a pure line, and certainly untrue in a mixed stock. Different stocks of mice differ in their susceptibility to cancer. They certainly differ in the median ages of death from cancer, defined as above. They also differ in the incidence of cancer at different sites. It is at least possible that the median ages of death from cancer at various sites are sufficient to characterize the differences between cancer incidence in different lines.

I have lately had to deal with a problem of this type in quite small groups of fish with which my wife has worked. When females are isolated, a few of them become hermaphrodites, and produce offspring without the aid of a male. And probably most of those which live for a year extrude unfertilized eggs. These events are recorded. But how are we to compare their frequencies in two groups each consisting of about sixty fish, in which the median ages of death are decidedly different?

The statistical methods which I have tried all involve dubious biological assumptions. Even if they did not, I do not see how to use the standard methods of comparison such as Student's test. Perhaps others may succeed where I have so far failed.

Any attempt to obtain life tables for a natural population of animals is extraordinarily difficult. It is possible for a natural population of plants, at least after the seeds have germinated and the cotyledons (the first leaves) are visible. But it would be extremely difficult; as in nature seedlings are commonly hidden by competing plants. The mortality must be immense, for some adult plants may produce an average of 100,000 or so seeds of which only one survives. But at least plants keep still. The best animal life table is that of Varley (1947) on a group of insects which stayed still because they spent a year in galls which they had made on plants. Varley opened about 50,000 galls, and concluded that about 1% of the larvae survived to emerge as adult flies, which then lived about a week. He enumerated and measured a dozen or so different causes of death in different parts of the life cycle.

Apart from this the best data are on birds. Fortunately most birds are fully grown before they leave the nest, so one can put numbered metal rings on their legs, and count those which are picked up after a given number of years. The fraction returned in Britain varies from under 1% for some small birds to about 15% for large ones. Most birds die in their first year of life, as they must, since each pair of small birds lay from 3 to 10 eggs, and the survivors start breeding after a year. After this a remarkably constant fraction dies each year. The fraction is about a half for birds the size of the mynah. The data so far show no signs of old age. On the other hand, data on wild sheep, which register their age by adding a new ring to their horns each year, show a great increase in mortality after about twelve years.

Sample surveys are essential when populations are very large. Let me give an example. Salt and others (1948) examined 20 cores from the same English pasture in November. They contained 42,753 arthropods, of which the majority were mites; and some were missed. There were 15,170 insects, but the numbers in a core varied from 250 to 1668, the mean being  $758.5 \pm 56$ . The density was  $2.6 \times 10^{11}$  per square kilometer, or  $2.6 \times 10^8$  per hectare. This is the approximate number of human beings. In May the number was only  $1.1 \times 10^8$  per hectare. The results are interesting as showing the extreme fluctuations in the numbers of some of the rarer groups of insects. Raw (1956) studied the Protura (primitive wingless insects) in the same area in a larger number of cores. The logarithms of  $n+1$  (adding 1 because of zero counts) were fairly symmetrically distributed, and 59% lay between  $\pm 0.3$  of the mean.

The samples are usually made by mechanical methods. But biological methods can be used. Lack and Owen (*J. An. Ecol.* 25) analysed 24 meals given by swifts (an insect-eating bird) to their young, by gently pressing the throats of the babies after the mother had left them. They found an average of 680 insects per meal. These had taken about an hour to collect. The frequencies of different species were significantly different from those collected at the same time in the same area by a mechanical method, but species common in one set were usually common in the other.

## THE SCOPE OF BIOLOGICAL STATISTICS

In the case of plants, or of animals which do not move quickly, at least during a part of their life cycle (like birds which are very immobile as eggs) one can study their distribution in a series of unit areas within a larger unit area, for example, wire squares with sides of one metre, thrown down at random. We may find a Poisson distribution. As you know, the second factorial cumulant of a Poisson distribution is zero. If its estimate is significantly positive we may say there is a tendency to bunch. This could be due to the patchy character of the soil, the sociability of the animals in question, and so on. Where the second factorial cumulant is significantly negative we suspect some kind of antagonism between individuals or mated pairs, in some cases amounting to a territorial system. Thus in England we can divide birds into those whose nests are often close together, for example, rooks, and those where each male defends a territory round his nest, for example most song birds. In the later case the mean number of nests in a unit area is less than its variance. This is one of the simplest statistical measures of animal social behaviour.

My own work has mostly been of a rather different kind. I have worked out the consequences of various hypotheses to see how selection, inbreeding, and so on, could be expected to act on populations. In the language of probability theory I have examined a number of Markoff chains. In few cases have observational results been exact enough to confirm my conclusions more than quite roughly. However, the work was necessary in order to arrive at a logically coherent theory of evolution.

I think that the time has now come when one can see some of the observational data which are needed. An important desideratum is a careful study of the interaction between neighbouring plants. And since rice is planted out much more regularly than wheat, it offers excellent material for such work. The interaction may be highly competitive. Sakai (1957) has found that some kinds of rice greatly lower the yield of their neighbours, whether of the same race or a different one. This effect may be so great that by selecting in each generation for the highest yield one obtains plants with a lower yield than the original mixture. For highly competitive plants, though they do well when surrounded by less competitive ones, may depress one another's yield very considerably. On the other hand, Gustaffson (1953) has found, in the case of barley, that mixtures of certain races may give a higher yield than either race by itself. Such phenomena can, I think, only be studied statistically by the full use of the techniques developed by Fisher. Their explanation will call for biochemical rather than statistical research. But the facts still demand much further statistical investigation. Neither Sakai nor Gustaffson has made all the experiments possible with their own material.

The improvement of animal breeds is based on artificial selection. This is a relatively simple matter in the case of sheep where both sexes produce wool, and males can be selected for their performance. It is much more difficult in cattle bred for milk production or poultry for egg production, where only the performance of females can be measured, and yet improvement is mainly by selection of males, since only a small fraction of males is bred from. Ideally one would like to choose a bull who

has begotten say ten daughters whose mean milk yield was notably above the mean of their mothers' yield. To do so one might have to wait till the bull was ten years old. This is practicable when artificial insemination is used. But in ordinary breeding one must be content with choosing a bull on the record of his mother, his grandmothers, his aunts, and so on, and perhaps some sisters, or half sisters. Dickerson and Hazel (1944) examined this problem statistically, and concluded that progress is quicker if the bulls are merely chosen on this basis and not on the performance of their daughters. If one waits for progeny tests the steps in improvement are larger, but there are fewer of them in twenty years. This question can only be answered by statistical methods, and the answer given for cattle is not necessarily valid for goats or poultry.

The improvement of plants is a more complicated matter, as will be seen from a consideration of some of the commonest trees round the Indian Statistical Institute. Bananas (*Musa* spp.) are asexually propagated. The differences between members of the asexual progeny of a single plant are usually slight and rarely inherited. The bananas systematically grown in Bengal belong to quite a small number of clones.<sup>1</sup> It should be possible to determine the mean yields of these clones as a function of soil conditions, time of year, and so on, and plan planting accordingly. There is no reason to think that the mean yields of the different clones have anything like a normal distribution. So the analysis of variance into within-clone and between-clone components might be somewhat misleading. The mango is also asexually propagated, although grafting slightly complicates the problem. It is possible to make new clones, both of the banana and mango. But they cannot be produced by simple selection.

On the other hand the palms are propagated by sexual reproduction like most animals. In some species such as the *khajur*, *Phoenix sylvestris*, the sexes are separate. But so far as I know it has not even been determined which sex yields most *gur* or toddy. The coconut palm or *narikel*, *Cocos nucifera*, is a hermaphrodite like most plants. It could and should be improved. But little can be done until the father as well as the mother of each nut planted to produce a new tree is known. This knowledge can only be obtained by climbing the trees, first to protect the young female flowers from pollen, and then to pollinate them from a known tree as father. Only then will it be possible to discover the effects of self-fertilization and the regression of a tree's yield on those of its parents.

I mention these facts simply to emphasize that even where, as with the palms, no genetics at all are known, the statistician will be well advised to learn some elementary biology.

I will indicate a few further fields of biology where I think statistical methods are needed. They have already been applied to some extent to the study of animal behaviour, but mainly, if I may say so, at a rather low statistical level. Some behaviour consists in the very frequent repetition of the same act, or pair of contrary acts,

---

<sup>1</sup> In this context a clone means all the plants derived from a single seedling by asexual propagation.



for example of a fish rising for a breath of air, and a sitting bird flying away from its eggs and back to them. It has been usual only to record the frequency of the act, or the mean interval between acts, the variance of this interval being merely used to test the significance of the difference between means recorded under different circumstances. However other moments of the interval distribution are just as interesting. Moreau (1939) found that intervals of absence from a nest had a smaller coefficient of variation than intervals sitting on eggs. This is intelligible, as too long an absence could cool the eggs, perhaps dangerously. Spurway and I found that the intervals between successive ascents for air of *Anabas testudineus* had a nearly normal distribution with a coefficient of variation about 25%. By altering the gas breathed, we could lengthen or shorten the mean interval, leaving the coefficient of variation and the shape of the distribution unaltered. Or we could barely alter the mean, but increase the variance about forty times, and make the distribution highly skew positively.

The application of statistics to experimental animal psychology has largely been to learning. I believe that it will be much more fruitful, at least for the next few years, to study activities which are either unlearned or completely learned, so that the learning process does not alter the animal's behaviour during our observation. I should like to study animal technology, and in particular the construction of nests by birds and webs by spiders. Consider, for example, a weaver-bird's nest. It may be made mainly from one species of grass. What is the mean number of blades used, and its variance between nests? What is the mean length of blade used in a given nest, and its variance? Do these vary between nests? The results could be compared with data on human craftsmen.

I have said nothing about the application of statistics to human psychology. This seems to have developed in rather serious isolation from other statistics, and it is at least possible that a wholly fresh attack on it might be useful. Nor have I dealt with human medical statistics. The mathematical principles involved are seldom difficult. On the other hand, the data are biased for a variety of reasons. For example, the fraction of cases of a disease reported varies with class and occupation. This is obviously true in India, but it is also true in Britain. If coalminers lose more days through sickness than clerks, this is not wholly because coalmining is the less healthy occupation. It is partly because coalminers are incapacitated by an illness which would not incapacitate a clerk. Again old coalminers have a rather surprisingly low mortality. This is partly because those who suffer from illness in middle age leave the occupation, and are registered as shop assistants, caretakers, or members of some other not too strenuous occupation, when they die.

Until the number of practising physicians in India is increased at least five times it is hopeless to expect adequate medical statistics in India. It is more important to save lives than to keep detailed records of death and illness. And there are not enough men and women to do both. It is not impossible that valuable information might be gained by Sample Surveys, but their planning would demand

much knowledge of social and economic facts as well as of statistical theory. I have only been able to help medical statisticians when they were very sure of their data.

In the diagnosis of disease, especially chronic disease, physicians are apt to rely on deviations from the median in one or more quantitative physiological characters. The mouth temperature and blood glucose concentration are well-known examples. However, the distributions of some such measurable characters are likely to be rather different in India and Europe, certainly on account of differences in climate and diet, and possibly on account of racial differences. For example, on going to a hot climate, Europeans usually increase their blood volume by 10% or so, thus diluting their haemoglobin. The blood has many functions. One is cooling the body by moving quickly through the skin. For this function a high volume and low viscosity are useful. I should thus expect healthy people in India to have a rather larger blood volume, with less haemoglobin and fewer corpuscles per millilitre, than comparable healthy Europeans. If so the criterion for a diagnosis of anaemia should be more strict than in Europe. We may hope that the research organized by Dr. B. C. Das in this Institute will furnish data on a number of measurable physiological and biochemical characters which will provide better norms for Indians than exist at present.

Statistical palaeontology is becoming important for two very different reasons. On the one hand, it gives us almost the only method of measuring rates of evolution directly. I can perhaps claim to have opened this particular field (Haldane 1944) by obtaining very rough measures of the rates at which the means of measurable characters changed in well established evolutionary lines such as the ancestors of horses. Mean rates were of the order of  $3\frac{1}{2}$ % per million years. On the other hand the means of certain characters in fossil molluscs are used to characterize particular British coal seams. The materials available are, however, biased in several ways. Smaller and thinner bones and shells are more likely to be destroyed, or to be thrown away by collectors, who will on the other hand pick out grossly abnormal specimens even if they are damaged. A statistician who wishes to work on such data should have at least some experience of field work in palaeontology.

I have said little of the mathematical problems raised. Pearson's and Fisher's researches both arose from the consideration of biological questions. But they have not exhausted the mathematical side of the subject. The question of experimental design has only been solved by Fisher for an important but limited class of experiments. Fisher's technique is based on the fact that agricultural soil is very variable. It is possible to standardize soil and other conditions in a green-house. Went (1953) has reduced the coefficient of variation of the height of individual plants in a pure line to 4%. Where such standardization is possible the question of experimental design is essentially one of costing. Standardization is expensive, but if it reduces the number of plants needed by 95%, it may be worth while, provided we are answering a biological question, and not trying to answer an economic one. For cultivators are certainly not going to keep rice plants in individual pots with controlled humidity, lighting and so forth.

## THE SCOPE OF BIOLOGICAL STATISTICS

The work of Bonnier and his colleagues (e.g. 1948) in Sweden gives a beautiful example of this kind of planning. He wished to work on the influence of diet, in childhood and later, on the milk yield of cows. Is it economical to feed them very well as calves, or can they be left to find their own food, and only pampered during pregnancy and lactation? He found that the mean square difference between milk yields of pairs of monozygotic twins was about five percent of that between calves taken at random from the same herds. He therefore collected such twin pairs from all over Sweden, at considerable expense, and brought them up side by side in his cattle sheds on different diets. This reduced his expenditure on food, shelter, and attendance by about 95%, which far more than balanced the cost of advertising, transport, and tests on blood and other characters of the twin pairs which were needed to establish that they were monozygotic. It should never be forgotten that the planning of biological experiments, like that of sample surveys, is an economic matter.

An unsolved mathematical problem is the statistical treatment of growth curves. Human male heights, for example, increase rather suddenly during a year or two at the age of puberty. But this age is variable. So the graph of the average height of 100 boys shows a slow rise spread over many years, and is not typical of any individual. The problem of finding a representative growth curve, and estimating the variation round it, is difficult. It would be interesting, however, to collect even twenty such curves in India for comparison with European and American data.

I am at present engaged in working out the theory of estimating rates of increase of populations with known birth rates. The rate at which a population will increase exponentially depends mainly on two parameters, the mean number of daughters per mother, and the mean maternal age at the birth of a daughter. The other moments of the maternal age distribution are less important, though they cannot be neglected. However, all these statistics are correlated when individual mothers are considered, and it is hard to give even the approximate sampling variance of the rate estimate, and would be very hard indeed to give its exact distribution, since the statistics for individual mothers do not have normal distributions.

I hope I have shown you that the field of biological statistics is wide open. Professor Rao has improved and standardized methods in what may be called the field of classical or Pearsonian biometry. But when we leave such data as measurements on adult human beings, distributions may be very far from normal. Salisbury (1942) gives figures for the numbers of seeds or fruits borne by different plants which are about as positively skew as the distribution of real incomes in India. That is to say, even the distribution of their logarithms is still positively skewed. For nobody could live on one tenth of the geometrical mean income, and plenty of people have more than ten times this income. It is a striking fact that if two or more variates are fairly highly correlated, for example length, breadth and depth in fish of the same species but different sizes, one can apply much correlation technique even if all the

distributions are extremely skew. For the breadths and depths of fish of a given length may have fairly normal distributions. Such methods are likely to be particularly useful in statistical palaeontology.

I hope I have shown you that the field of biological statistics is very broad, and that there is still plenty of work to be done in it, both practical and theoretical. I would only warn you that practical work will probably be inefficient without some theoretical knowledge, and theoretical work may be ornamental but will not be useful without some practical experience. It should be one of the tasks of this Institute to see that theory and practice are not too widely separated.

## REFERENCES

- ATTFIELD, M. (1951) : Inherited macrocytic anaemias in the house mouse. III. Blood-cell diameters. *J. Genet.*, 50, 250.
- BONNIER, G., HANSSON, A., AND SELJEVOLD, H. (1948) : Studies on monozygous cattle twins. IX. The interplay of heredity and environment on growth and yield. *Acta. Agric. Suec.*, 8, 1.
- CALLAN, G., AND SPURWAY, H. (1951) : A study of meiosis in interracial hybrids of the newt, *Triturus cristatus*. *J. Genet.*, 50, 235.
- DICKERSON, G. E., AND HAZEL, L. N. (1944) : The effectiveness of selection on progeny performance as a supplement to earlier culling in livestock. *J. Agric. Res.*, 69, 469.
- GUSTAFSSON, A. (1953) : The cooperation of genotypes in barley. *Hereditas*, 39, 1-18.
- HALDANE, J. B. S. (1949) : Suggestions as to quantitative measurement of rates of evolution. *Evolution*, 3, 51.
- (1956) : The detection of autosomal lethals in mice induced by mutagenic agents. *J. Genet.*, 54, 327.
- LACK, D., AND OWEN, B. F. (1955) : The food of the Swift. *J. An. Ecol.*, 24, 120-136.
- MORBEAU, R. E. (1939) : Numerical data on African birds' behaviour at the nest : *Hirundo s. Smithii* Leach, the Wiretailed Swallow. *Proc. Zool. Soc.*, 109A, 109-125.
- RAW, F. (1956) : The abundance and distribution of protura in grassland. *J. An. Ecol.*, 25, 15-21.
- RENDEL, J. M. (1943) : Variations in the weights of hatched and unhatched ducks' eggs. *Biometrika*, 33, 48.
- SALISBURY, E. J. (1942) : *The Reproductive Capacity of Plants*. London, Bell.
- SAZS, G., HOLLICK, F. S. I., RAW, F., AND BRIAN, M. V. (1948) : The arthropod population of pasture soil. *J. An. Ecol.*, 17, 189-190.
- SAKAI, K. (1957) : Studies on competition in plants. VII. Effect on competition of varying numbers of competing and non-competing individuals. *J. Genet.*, 55, 227.
- VARLEY, G. C. (1947) : Natural control of population balance in the knapweed gallfly (*Urophora jaceae*). *J. An. Ecol.*, 16, 139.
- WENT, F. W. (1953) : Gene action in relation to growth and development. I. Phenotypic variability. *Proc. Nat. Ac. Sci.*, Washington, 39, 839.

Paper received : July, 1958.