THE HOME UNIVERSITY LIBRARY
OF MODERN KNOWLEDGE

CLVI

STATISTICS

STATISTICS

By

L. H. C. TIPPETT

STATISTICIAN TO THE BRITISH COTTON
INDUSTRY RESEARCH ASSOCIATION

OXFORD UNIVERSITY PRESS

LONDON NEW YORK TORONTC

First published in 1943 and reprinted in 1944

PRINTED IN GREAT BRITAIN

CONTENTS

CHAPTER I

INTRODUCTION

STATISTICS affects everybody, and touches life at many points. As citizens we help to provide statistical information—our very entry into the world and exit from it are recorded statistically—and propagandists daily try to convince us of something, or even to fool us, by means of statistical facts and arguments. The running of the community through its institutions of government and business depends very much on statistical information, and this dependence increases as business tends to become concentrated in larger concerns and the government intervenes more and more to plan our economic and social life.

The propagandists, administrators, and business executives who use (and misuse) statistics are fairly numerous; and to them may be added such people as politicians, social students, and social reformers who employ statistical facts and methods to provide a basis for policy. Such facts and methods also have an important place in the development of sociology and economics as sciences; the methods are very important to experimentalists in most branches of biology, and are used by workers in the more exact sciences of physics, chemistry, and engineering. Statistical ideas are at the root of many current theories in biology, physics, and chemistry: indeed, a statistical approach is probably one of the most characteristic features of modern science. Finally, statistics as a subject is naturally a major interest to the comparatively small body of professional statisticians.

As a result of the many approaches to the subject, the word *statistics* and its associated words *statistical* and *statistician* have various meanings. First we have the dictionary definitions, in which *statistics* refers in the singular to the subject as a whole, and in the plural to numerical data. I shall adopt both usages. To the 'man in the street' statistics are just figures, and he is inclined to think of the statistician as being primarily one who counts the numbers of things. To the economist, used to the qualitative ideas of economic theory, *statistical* is almost synonymous with *quantitative*. To the physicist, *statistical* is the opposite of *individualistic* or *exact*, since to him statistics is a subject that deals, above all things, with groups and probabilities rather than with simple entities and certainties. To the experimental scientist who is used to gaining knowledge by conducting experiments under controlled conditions, statistical methods are those which are employed when accurate experimental control is impracticable or impossible. The field of application of statistics is mostly (but by no means entirely) economic, and so the statistician is sometimes thought of as a kind of economist. On the other hand, statistical methods are basically mathematical, and many people think of a statistician as something of a mathematician. One might almost say that the mathematician accepts the statistician as an economist, and the economist accepts him as a mathematician. Some cynics think statistical methods so uncritical that one can 'prove' anything by them; and others think they are so critical that they can prove nothing. At the other extreme are those enthusiasts who think that as a means of increasing knowledge the power of statistics is boundless and almost magical. These views are justifiable but incomplete; and the purpose of this book is to give a complete and (as far as I ca. balanced view of the whole subject.

There are several general considerations which n.;th profitably be borne in mind when approaching th. subject of statistics.

First, it is both a science and an art. It is a science in that its methods are basically systematic and have general application; and an art in that their successful application depends to a considerable degree on the skill and special experience of the statistician, and on his knowledge of the field of application, e.g. economics. Statistical methods are not a kind of automatic machine into which numbers can be put and from which perfect results can be taken. Nevertheless, the subject is not a closed mystery, and I believe that it is not necessary to be a statistician to appreciate the general principles underlying it.

As a science, the statistical method is a part of the general scientific method, and is based on the same fundamental ideas and processes. This point will frequently come up in this book, and suggests one reason why the study of statistics is good educationally. It teaches the scientific method in terms of things of everyday experience, and inculcates a habit of scientific approach to ordinary economic, social, and political problems. It will be seen, however, that statistical methods have their own special features. These arise from the fact that the data are not simple, like those that usually result from a well-designed and well-controlled scientific experiment, but are relatively complex, being the result of a number of causes all operating together without control. Statistics deals with figures that are subject to uncontrolled variation.

the word *st* ... ıt statistics has in common with
and *statist* ... ts is that it is not finished and
the dict ... developing. Despite its powei
the si ... ss, it has limitations and im-
to pay ... developments will undoubtedly

...ects included under statistics
, .. any, statisticians are expert in all
...ches. Some specialize in the development of the mathematical theory underlying statistical methods, and are essentially mathematicians. Others are interested in the methods themselves, both elementary and advanced, and in their general application to almost any field, although they often have also some special experience of one field. There are also statisticians who are able to use with confidence only elementary methods—perhaps fairly simple tables, diagrams, and averages—but who have a very wide and deep knowledge of some field of application. As will be seen later, knowledge of this kind is very important, and its use makes the work of such a statistician much more than the mere clerical work of tabulation that it sometimes seems to be. Statisticians in this category specialize; and often are as much economists or sociologists, say, as statisticians. One who is an expert in trade statistics may or may not know much about the statistics of public finance, but he will probably know very little of vital statistics; and an expert in vital statistics who wishes to deal with agricultural statistics, say, may have much to learn and much experience to gain.

CHAPTER II

THE RAW MATERIAL

THE conception of statistics as having to do with figures is the most popular one for very good reasons. Many of the questions that are the subject of common conversation and controversy require numerical data for their resolution. Trains are more (or less) crowded than buses; the English are better (or worse) patrons of sport than of the arts; women drivers are more (or less) competent than men drivers; and so on. These are the kinds of questions that are argued in newspaper columns, drawing-rooms, common-rooms, and public-houses. The disputants give their various experiences both relevant and irrelevant; one always travels by train and never has to stand; another once had to stand on a train journey but has never stood in the bus; another has his own motor-car and does not use buses or trains, but knows that road transport gives a better (or worse) goods service than the railways; and so the discussion proceeds, and probably leaves everyone at the end with the same opinion as that with which he started. But if someone intervenes with really reliable and cogent numerical information, based on a wide experience, the question is settled and the discussion 'peters out'. Life would be dull if we were debarred from discussing questions of fact we do not fully understand, and the 'know-all' who regularly ruins such discussions with his facts is little more than a bore. But we cannot afford to trifle with important subjects by ill-informed controversy.

General impressions are entirely untrustworthy, for some facts and events strike the imagination more than others and are more easily remembered. For example, if we have a theory that the weather changes with the rising and setting of the moon, we notice the occasions when our theory is borne out and forget those when it is not; and it is only by taking systematic records that we can arrive at the truth of the matter. Even when a general impression is qualitatively correct, it may be quantitatively incorrect. If the morning train to town is occasionally late, we are apt to feel that it is more often late than not, whereas an actual count might show that it is late on only one morning in ten, on the average. The following instance of a general impression being corrected by numerical data exemplifies a common occurrence. A few years ago the L.M.S. railway company analysed their passenger statistics to show how many journeys of various lengths had been made, and Mr. Ashton Davies reported: 'A great deal of valuable and interesting information came to light, much of it contrary to the then current opinion. For example, it was quite a common belief that the railways had lost most of their former short-distance passenger business, owing to the competition of road transport. The dissection, however, revealed that quite a significant proportion of the passenger business of the L.M.S. Company still consisted of really short-distance traffic.'

We shall see in later chapters that statistical methods are applied to the results of physical, chemical, and biological experiments and observations, as well as to results obtained in social and economic investigations. The making of observations is usually a major part of a research in an experimental science, and instruction in experimental technique and in the handling of the necessary apparatus forms a large part of the training of the physicist, chemist, and biologist. This is no place for a discussion on such a subject, which belongs properly to books that specialize on the sciences concerned. In social and economic research, on the other hand, the collection of the material requires little apparatus beyond a pen or pencil and paper, and no experimental technique; and possibly because of this the subject does not always get the attention it requires. Considerable knowledge and experience are necessary to know where to go for statistical material, how to ensure its accuracy, and how to interpret its meaning; and since these are matters that are regarded as falling within the scope of statistics, we must consider them at some length.

In order to obtain statistical material we may either go to the records of some public body that collects and publishes statistics as a routine, or make a special survey.

The most important routine collectors and suppliers of statistics are governments. The systematic recording of trade statistics by the English Government started with the appointment of an Inspector General of Exports and Imports in 1696; the first British census of the population was held in 1801. Since the early years of the nineteenth century the volume and scope of British official statistics have increased enormously and continuously right up to the present day.

The statistical publications of the British Government are indexed and briefly described in a Guide to Current Official Statistics, published annually by the Stationery Office. There are listed in this Guide some 500 reports, issued by every department of state, and covering a wide range of the nation's life and activities. The most important and well-known subjects are finance, trade and industrial production, employment and unemployment, prices, health and mortality, and population; but there are also more romantic subjects, such as reports of coroners' proceedings on treasure trove and statistics of smuggling seizures.

Government statistics were originally required for administrative purposes, and their publication to and use by independent investigators is in some degree a by-product. Nevertheless, some government departments, at least, take seriously their function as collectors and suppliers of statistics for general use, and it is well understood that one of the chief values of published reports is to provide material for independent investigators. The very existence of the Guide suggests this. Some government officials who deal with statistics are also Fellows of the Royal Statistical Society, where they have contact with statisticians outside the government service. These officials are ready to consider the suggestions and criticisms of statisticians and to be somewhat influenced by them in deciding the form and content of published returns; and any investigator who establishes his *bona fides* can often (not always) obtain access to unpublished statistical details.

Much good can be said of British official statistics, but considerable improvements are possible. There are substantial lacunae in the subjects dealt with. The population census gives much information about the social conditions of the people, but the collection of figures about the important factor of wages is left largely to unofficial investigators. Moreover, a census taken decennially is too infrequent and there is a strong demand for a quinquennial census. Statisticians and business men both want a Census of Distribution to give information, now almost entirely lacking, about the distributive industry.

A second complaint is that too long an interval sometimes elapses between the collection of material and its publication. For example, the final report of the Census of Production of 1930 was published in 1935. A third complaint is that statistics about similar or related subjects, collected by different government departments, are not co-ordinated, so that they often differ in scope and definition. This makes it impossible to use these data for making comparisons and considerably reduces their usefulness. The volume and cost of official statistics are so great, their subjects so comprehensive, the agencies that use them so many, and the purposes for which they are used so varied, that there seems to be an unanswerable case for a central department to secure the efficient collection and publication of official statistics and avoid inconsistencies, overlapping, and waste. However, the case has not yet been conceded by the British Government.

The governments of other countries also publish official statistics, varying in comprehensiveness and reliability, and the League of Nations is responsible for a considerable amount of important statistical work.

In Great Britain, many semi-public bodies such as the Bank of England, the stock and produce exchanges, and trading associations, regularly publish statistical material, mostly of a commercial character; and much of this is reproduced in financial journals and in the 'city' columns of daily newspapers.

Statistical information is given in a 'potted' form in a variety of year-books; for example, the Statistical

Abstract for the United Kingdom, published annually by the Stationery Office, contains about 400 pages of tables summarizing British official statistics. Such summaries have their uses, but also their dangers, for the tables are given with very little explanation and there is always a risk that the reader may misinterpret the figures. In addition, statistical and economic journals, notably the Journal of the Royal Statistical Society, contain many critically prepared digests of statistical information concerning a variety of subjects, and although these do not rank as primary sources, they can often be used by the inexpert with greater confidence than the original unedited figures.

Special surveys for obtaining statistical information are made by governments, unofficial bodies, and private individuals; and in this field the last two have led the way. Indeed the chief function of unofficial surveys has been to supply the deficiencies of official statistics. The Manchester Statistical Society and the Statistical Society of London (now the Royal Statistical Society) were founded in 1833 and 1834 with declared objects that included the collection of statistics 'illustrative of the condition of society', and in the early years of these societies this formed a fair proportion of their activities. In 1886, Charles Booth, a London shipowner and merchant, started his famous and very extensive survey of the conditions of life of the people of London. Because of its comprehensiveness and its statistical character this is regarded as a pioneer work. It has been followed by a very large number of social surveys in different parts of this country, and even more have been carried out in the U.S.A. I think it is true to say that, until recently, most non-official statistical activity has been concerned with the social condition of the people. Special statistical inquiries are also made in the commercial and political worlds.

The complicated analysis to which statistical data are often subjected, and the highly condensed form in which they are summarized as averages, give the final results of an investigation a form and order that often are not obvious in the original figures, and an appearance of accuracy and precision they do not necessarily possess. In contemplating the finished work it is all too easy to forget the raw material from which it is made. Nevertheless, no statistical results can be reached that are not already implicit in the data, and the accuracy of the former depends on that of the latter. It therefore behoves anyone who uses statistics to exercise care in obtaining them, and if he is going to use those already published, to examine them carefully for errors and to understand their exact meaning.

In order to do this effectively much knowledge is needed of the way in which the figures are collected, of the circumstances surrounding the facts recorded, and of the kinds of errors that can arise. As a check on accuracy, the results of one inquiry can often be compared with those of another to see that they are reasonably consistent, and the data can also be tested for internal consistency. For example, when vital statistics were first collected in some colonial dependencies in Africa, some twenty years ago, it was found in one instance that more children died under the age of one year than were born. That is an extreme example of internal inconsistency exposing error! The following quotation from Mr. B. Seebohm Rowntree's

1941 report on his social survey of York gives a picture of the care and attention which are devoted to this business of the collection of reliable statistics: 'Obviously, in making a house-to-house inquiry everything depends upon the skill, tact, and reliability of the investigators. It took some time to discover just the right people, but eventually seven were found, five women and two men, on whose work full reliance could be placed. . . . Moreover, a number of "check" visits were paid, at random, or to cases that seemed abnormal, and in that way the accuracy of the returns was tested and verified.'

In the following paragraphs I give a few examples of the difficulties and pitfalls that exist in the collection and interpretation of statistical data.

It is obviously foolish to place reliance on figures that are palpably false. A woman's statement of her age is proverbially unreliable, and the information gained by asking questions on matters that are ill-defined, or matters of opinion, depends somewhat on the way the questions are framed. Market investigators have found, when investigating the reasons why people buy particular brands of goods, that the direct question is not likely to give the true reason, as people do not all indulge in honest self-examination, and those who do may not be honest with the investigator. Indeed, the person questioned may not have thought about the subject before, so that the result of the inquiry is influenced by the inquiry itself. This must often happen in psychological investigations. Reliable data on household expenditure are always difficult to obtain—how many housekeepers keep even approximate records of the way in which they distribute their expenditure?

It may have been noticed by some readers that government officials in collecting returns often show what seems to be almost a passion for putting people into classes, unless the data can be given in a well-defined form such as age or place of birth. When registering for military service, for example, each man has a classification number describing his occupation, and however unique that occupation may be it must be fitted into a class. This is because classification is a fundamental part of the statistical method, as we shall see later, but it is done by the official 'on the spot' because only there is the complete information available which enables an accurate assignment to the appropriate class to be made. Difficulties often arise because of 'border line cases'. In the Census of 1931 householders were required to give the industry in which the members of their household were employed. To what industry belongs a man who does odd jobs? And if an unemployed man from one industry does some other temporary work, what time must elapse before he is regarded as having left the original industry for good and all?

Uniformity and accuracy can be attained in such instances only if very full and precise definitions are given—arbitrarily if necessary. Such definitions sometimes lead to amusing results. For example, in the British Census, a street singer is a musician, an organ grinder a worker in Sundry Industries, and the occupation of a pavement artist is not specified. Statisticians prefer data that are precise, even if apparently a little foolish, to data that are vague. The Metropolitan Police once had a book for recording goods 'suspected stolen', but the instructions as to what to enter in this book were not precise, and there was a tendency to include actual thefts; it was feared that this tendency varied from one district to another and from one time to another, thus destroying the value of the returns. When, in 1932, this book was abolished, and the police had to make up their minds whether or not the goods were stolen, the number of recorded indictable offences in the Metropolitan Police area (not necessarily the number of crimes or indictments) rose from 26,000 in 1931 to 83,000 in 1932.

The reliability of statistical observations depends very much on the way in which they are made as well as on the ease with which the required information can be given. Much information is gathered from returns and questionnaires completed by people who are not interested in statistics—citizens, taxpayers, business men, farmers, factory managers, public officials, and so on—and it must be recognized that people do not like filling in forms. The farmer is interested in growing and selling crops, and he regards the making of statistical returns as a pestiferous waste of time; even a statistician would probably be impatient if, for the information of another statistician unknown to him, he had to interrupt his work on, say, the world trade in ants' eggs, to make a return of the number of man-hours occupied in the investigation. Therefore it is wise not to rely too much on the conscientiousness of people in completing returns; and to remember that the results are likely to be reasonably reliable only if the questions are few, straightforward, and easy to answer.

Where the required information is complicated or difficult, enumerators or 'field-workers' are usually employed. The enumerators employed in making the

Census have very full instructions as to how the census forms should be filled up, and experienced field-workers such as usually conduct social surveys know the snags and are not likely to be misled, so that the figures they obtain are usually fairly accurate.

Although data made up of defined measurements are preferable, the statistician often has to deal with vague quantities that are matters of personal judgement, such as general health and intelligence. To give estimates of such things any value at all, the observer must be specially careful to standardize and define the basis of his judgements as far as possible, so that he can obtain consistent results that may at least be valid for making comparisons. One important stage in doing this is to divide the quantity into a number of parts, and to give points for the parts separately, adding the points to obtain the final result. School examinations provide an example. The candidate answers a number of questions, each of which is marked separately; and some examiners even subdivide the marks for a question, giving so many for the correctness of the answer, so many for the correctness of the method by which the answer is obtained, so many for the orderly presentation of the argument, and so on. Recent investigations have shown that examinations do not measure attainment with great exactness—a result which shows that even when considerable care is taken, it is difficult to make reliable data that have a subjective basis.

However accurate and self-consistent statistical results may be, they cannot be used safely unless everything is known of the way in which they were obtained and of the real meaning behind the figures. It is not often that all the detail surrounding any body of information is published with the figures, and so things are not always what they seem. For example, British criminal statistics give, not the numbers of crimes committed, but the numbers reported to the police; and the two are very different. The proportion of crimes that become known to the police varies with the kind of crime and from time to time, according to the changing attitude of public opinion to the various crimes; and the statistics give very misleading impressions of the amount of crime extant.

I have already stated that some of the categories used in describing data may have to be defined arbitrarily. It is necessary to be aware of differences that exist between departments of one government, and between countries, in the definitions they adopt for what is nominally the same quantity. Those who use statistics in the form of series extending over some time have also to be on their guard lest some change in definition or other basis should break the continuity of the series. For example, the figures published periodically of the numbers of the unemployed include only insured workers registered at the exchanges on certain dates, and these are affected from time to time by legislative changes in the classes of workers who may be insured (e.g. in the age limits) and in the qualifications for unemployment benefit. Statistics of causes of death extending over long periods of time are apt to be affected by changes in medical knowledge and (dare a layman suggest?) fashion causing changes in diagnosis. So important is continuity in recorded statistics that statisticians almost prefer an existing unsatisfactory basis to be maintained rather than suffer changes that may in many ways be improvements; and they are very insistent that when changes are made, two sets of figures should be obtained for some time, one on the old basis and the other on the new, so that the old series can be joined to the new.

The intelligent interpretation of final statistical results often requires a knowledge of unrecorded circumstances surrounding the events recorded. For example, in an investigation on accidents in naval dockyards it was found that the *recorded* accident rate among apprentices tended to decrease year by year through their apprenticeship, whereas that among naval artificers tended to increase, although both groups were doing similar work. The explanation of this difference is that the apprentices worked under industrial conditions and lost 'time' and money when away from work because of an accident; the artificers worked under service conditions and did not suffer this loss.

The method of inquiry by sample, which is much used in social work, has its own special difficulties and sources of error; that method will be dealt with in Chapter VI.

To give readers a more concrete and integrated picture of what is involved in the collection of data for a statistical inquiry, I am going here to give some detailed comments on an actual example. The Ministry of Transport has published three Reports on Road Accidents occurring in Great Britain in the years 1933, 1935, and 1937, and these contain statistical summaries of a number of details of the accidents; the inquiries that provided the information for these reports are the example. In commenting on the collection of the data, I know nothing of what went on 'behind the scenes', but some information is given in the Reports and the rest has been surmised.

Presumably the Ministry hoped that from a statistical summary of the circumstances surrounding road accidents, something would be learnt of the causes, and this purpose, as well as administrative and practical considerations, would be borne in mind when deciding what data to collect. The police in various parts of the country reported the details of the accidents and it was therefore necessary to ask only for such information as such a scattered body of men could give reliably and uniformly. In the earlier Reports, estimates were given of the speeds of the vehicles just prior to the accident, but such estimates were admittedly unreliable and were not given in the 1937 Report. There are also other details that would have been very useful but in the circumstances had to be omitted. Thus, the previous accident and medical history of the drivers would have helped to determine whether medical or psychological fitness had anything to do with the tendency of a driver to become involved in accidents.

Since it is not to be expected that the police would fill in the necessary questionnaires with as much zeal as they give to their ordinary duties, it was desirable to limit the number of questions and not to ask for unnecessary details. The statistician who organized the inquiry had to know enough about traffic conditions to decide which questions were important, and which could be omitted as having little or no importance (e.g. the colour of the vehicle or of the upholstery).

Each accident was reported on a form, thus ensuring that no details were overlooked and the data were collected uniformly and systematically by the various police officers. Some of the details were definite and required little explanation, e.g. the date, place, and time of the accident and the number of persons killed and injured, and so on. Others, such as the cause of the accident, needed careful definition. The police were not asked to record the cause in their own words but sixty-four possible causes were listed and fully described, and the policeman who reported the accident stated which one of these operated at the accident in question. Thus a basis was provided for grouping the accidents according to cause. Those who described these causes needed a considerable knowledge of road traffic to ensure that the list was exhaustive. Another piece of information asked for was the extent of the injuries to injured persons—whether fatal, serious, or slight, and elaborate instructions were also given for defining this.

The development of the details of an investigation of this kind requires much thought and planning, and only a superman would be able to decide the best methods straight away. It is quite clear, from the changes between the 1933 and the 1937 inquiries in the information sought, that the Ministry officials gained experience as they went along; the 1937 questionnaire and method of inquiry, which are given and described in the corresponding Report, are a result of this experience, and are a good example of the first stage of a well-arranged statistical investigation.

Accuracy and reliability in the data are important because their lack cannot be supplied by elaboration and care in the subsequent statistical treatment. However, as this is an imperfect world, most data are imperfect in some degree and many are very imperfect. Nevertheless, they usually contain some information and so are far from valueless; and it is the statistician's job to make the best he can of them. It is a mistake to think, as some do, that inaccurate or unreliable figures should not be given careful treatment; they may not merit it, but they certainly need it. For example, extra care is necessary to allow for the inaccuracies and avoid arriving at false conclusions. Thus, the statistician will first do all he can to obtain data that are as precise as possible, and will then apply his methods of analysis to make the best possible use of the figures he obtains.

CHAPTER III

ARRANGING AND PRESENTING THE MATERIAL

THE results of the first stage of a statistical inquiry are sometimes a few fairly simple figures which can easily be presented and understood without any special treatment; but more often there is an overwhelming mass of data and detail. The first task of the statistician is to reduce these in the two senses of (*a*) making less the amount of detail and (*b*) bringing the data into a form whereby the significant features stand out prominently. The statistician must get out of the situation in which he cannot see the wood for the trees. It is easy to state in general terms how this is done: the unimportant details are decided upon, and the data arranged so as to suppress these and leave the important features clearly expressed. The process is essentially one of summarizing.

In fact, this is already started when the field and scope of the inquiry are chosen. With the whole universe before him, the investigator chooses the subject of, say, the housing conditions of one city at one time, and he ignores as irrelevant everything except a few particular facts for the city, such as the numbers and sizes of houses, their distribution, the numbers and composition of the families living in them, and perhaps the incomes of the families. It is not supposed, of course, that the ignored facts are absolutely unimportant, indeed they may afterwards have to be considered even in relation to the original subject of the investigation. The state of housing in a city may, for instance, later be related to unemployment or to other things that are omitted from the original inquiry because it is impossible to deal with everything at once. But some selection must be made, although this may depend partly on such accidents as the investigator's interest. This process of isolating some small part of the universe for study is common to all scientific investigation.

The first and most important step in the statistical reduction of data is usually to group into one class the items that, for the particular purpose in view, need not be distinguished. When many items are put into several groups in this way they are classified. For example, the Statistical Abstract gives the yearly values and quantities exported from the United Kingdom of over a hundred kinds of articles. To reduce these figures to manageable proportions, and obtain a broad picture of the export trade, it may be sufficient to use the Board of Trade classification the three broad classes: I. Food, Drink, and Tob II. Raw Materials and Articles Mainly Unm factured; III. Articles Wholly or Mainly Manu tured. In this scheme no distinction is made betw biscuits and fish, both of which come in Group between coal and wool, which are in Group II; between coke, steel, and horses, all of which are Group III.

Sometimes the subject falls easily and naturally in a few categories. Thus, if families are grouped a cording to the number of children, the categories wi naturally be 0, 1, 2, 3, etc. children. Frequentl however, the subject is such that the classes have t be created more or less arbitrarily, as for the export just mentioned. There are three points to be observed in making such classifications.

First, any given body of results can usually be classified in many ways, and the best way will depend on the purposes of the inquiry. If the aim is to relate changes in export trade to changes in employment, the exports might be so grouped as to include coal among manufactured articles, because the coal industry employs a lot of labour. For other inquiries it might be better to group the articles according to industries —fish, coal, iron and steel, textiles, engineering, and so on, or according to the amount of shipping space required per million pounds' worth of the article. One difficulty in the way of using existing published igures is that they are not always classified in a way hat is suitable for the particular inquiry. The choice f the basis of classification is not a matter of statistical ethod, but requires special knowledge of the subject investigation.

The second point, implicit in the whole idea of classification, is that all items grouped together should, for the purposes of the particular inquiry, be sufficiently alike; that each class should be homogeneous. Table 1 gives a summary of the death rates in England and Wales in 1938, and in the upper part of the table all

TABLE 1

Deaths during 1938 of Persons of Various Ages, per 1000 Persons of Corresponding Age living at the Mid-Year. England and Wales

Age, years	*Death Rate*
0–10	8·4
10–20	1·5
20–30	2·6
30–40	3·1
40–50	5·7
50–60	12·7
60–70	29·2
70–80	73·3
80–	173·8
0–1	55·1
1–5	4·6
5–10	1·9

'e in any one decade of life are classed together. ïgures given in the lower part of the table show ıe 0–10 years group is far from homogeneous, for as the average death rate for the group is 8·4, ır the first year of life is 55·1, the rates for the ing years being very low. For general mortality , the grouping by decades may be sufficient, but udy of infantile mortality much finer grouping ssary; and figures are in f-separately for the first few months of life, although the death rates used in this connexion are not expressed as in Table 1, but as deaths per 1000 live births. The death rates for the separate years over 80 are also probably far from uniform, but we are not often interested in studying mortality at these ages and the variation is therefore unimportant and may usually be ignored.

Lack of homogeneity within the groups may always be detected by examining the figures, as we have done those in Table 1, but the investigator can often decide from his general knowledge what variation is likely to occur, and can adopt a scheme of classification accordingly. Usually, the variation in the whole material is so great and complex that it is impossible to have classes that are perfectly uniform. However, the statistician must classify, and so some variation within classes must usually be tolerated. Skill and experience are necessary to steer safely between the Scylla of having too few broad classes, with much variation within each, and the Charybdis of having too many fine classes, each with very little variation.

Consideration of the third general point about classification modifies in some degree the application of the second. Consider Table 2. According to the full classification given there it appears that pedestrians are the 'villains of the piece' in causing road accidents; of all the classes, they caused most accidents. But it is unfair to class all pedestrians together and to separate the various kinds of drivers. When the drivers are classed together it is seen that they caused 37·6 per cent. of the accidents and were almost as culpable as pedestrians. Thus we see that false impressions may be created if the classes are not of the same rank,

From classical mechanics to quantum mechanics

In Lieu of an Introduction

Atomic energy. Radioactive isotopes. Semiconductors. Elementary particles. Masers. Lasers. All quite familiar terms, yet the oldest is hardly twenty-five years of age. They are all children of twentieth-century physics.

In this age, knowledge is advancing at a fantastic rate, and every new step opens up fresh vistas. The old sciences are going through a second youth. Physics has pushed out ahead of all others and is pioneering into the unknown. As the front broadens, the attack slows up only to make renewed thrusts forward.

To get at the secrets of nature, physics has had to find powerful instruments, to devise precise and convincing experiments. At the physics headquarters are hundreds and thousands of theoreticians mapping out the offensive and studying the trophies captured in the experiments. This is no struggle in the dark. The field of battle is lighted up with powerful physical theories. The strongest searchlights of present-day physics are the theory of relativity and quantum mechanics.

Quantum mechanics came in with the twentieth century. Date of birth: December 17, 1900. It was on this day that the German physicist Max Planck reported to a meeting of the Berlin Academy of Sciences Physical Society on his attempt to overcome one of the difficulties of the theory of thermal radiation.

Difficulties are a common thing in science. Every day scientists come up against them. But Planck's encounter had a very special significance, for it foreshadowed the development of physics for many years to come.

An enormous tree of new knowledge has grown out of the seminal ideas expressed by Planck, which served as a starting point for amazing discoveries far beyond the imagination of the wildest science-fiction writers. Out of Planck's concepts grew quantum mechanics, which opened up an entirely new world – the world of the ultrasmall, of atoms, atomic nuclei and elementary particles.

The Outlines of the New World

But didn't people know anything about this atom before the twentieth century? In a way they did, that is, they had guessed and conjectured.

The inquisitive human mind had speculated upon these things and had long imagined what became a real thing only many centuries later.

In ancient times, long before the first travellers laid their paths of discovery, man had guessed that there were people and animals and land beyond the little area in which he lived.

In the same way, people felt that there existed a world of the ultrasmall long before it was actually discovered. One did not need to go far in search so to speak. It is impossible to say what is the correct grouping in such instances—probably none is absolutely correct—but we should realise that the grouping can affect the impression created by the data.

TABLE 2

Numbers of Fatal Road Accidents in Great Britain caused by Various Classes of Road Users, 1936–37

Class of Road User	*Accidents*			
	Number		*Percentage*	
Drivers of private motor-vehicles, except motor-cycles	868	2,201	14·8	37·6
Drivers of motor-cycles	868		14·8	
Drivers of public conveyances	92		1·6	
Drivers of vans, lorries, etc.	333		5·7	
Drivers of other vehicles, except pedal cyclists	40		0·7	
Pedal cyclists	1,051		17·9	
Pedestrians	2,470		42·2	
Other persons	133		2·3	
Total	5,855		100·0	

The most usual way of presenting statistical information is in a table of figures. Early in the nineteenth century there was some controversy between those who preferred to present results in a literary form and those who preferred tables and who were accused of presenting only the 'dry bones'. However, the 'table school' won the day, and it is perhaps an echo of this controversy that the original prospectus of the (then) London Statistical Society declared the aims of the Society 'to confine its attention rigorously to facts—and, as far as it may be found possible, to facts which can be stated numerically and arranged in tables.'

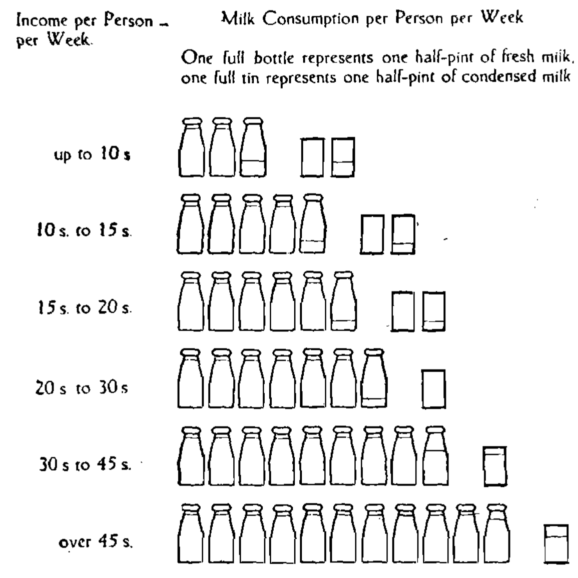
Statistical data are usually presented in tables, even in ordinary newspapers, but some writers still seem to prefer giving their figures in a more literary form. The question may be one of taste and training, but I find the tabular method of expression much clearer. It does not make the figures any less dry to have them strung out in sentences and joined by words and phrases such as 'whereas', 'on the other hand', 'as against', and so on. I am not here criticizing the use of words to point out special features or contrasts shown by results given in tables.

There is an art in arranging a table to present data economically and clearly, and in a way to facilitate any comparisons that the reader may be required to make. However, the arrangement cannot add to or subtract from the significance of the figures, and I need not write anything further on the subject here.

Diagrams and charts are also much used in presenting statistics, and have a value because even statistical ones give some delight to the eye and add a spark of interest to a paper. Their chief importance, however, is that they give a picture of the broad statistical facts that is more readily taken in than a table. It requires a careful examination of the figures of a table to appreciate their full significance, and great concentration of thought is necessary to keep the general picture in mind while reading the figures in detail. Magnitudes are more easily appreciated and remembered when conveyed to the mind by pictures than by numerical figures. On the other hand, the broad picture given by a diagram is not as exact in detail as that given by a table of figures, and since it is somewhat affected by the way in which the diagram is made, it may even give a misleading impression. I am now going to give a few examples to illustrate these

Figure 1.

Relation between Milk Consumption and Income.



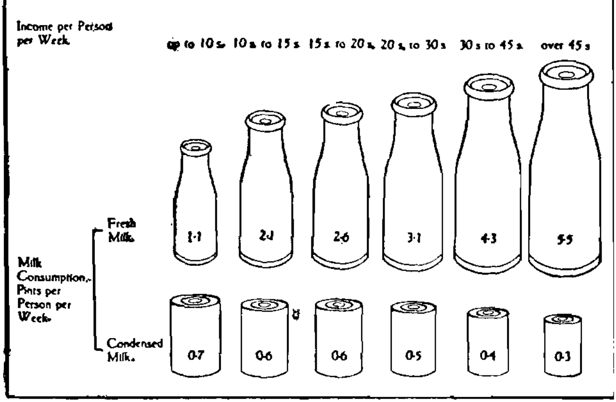
statements and to show how some important types of diagram are read.

Figure 1 shows Sir John Orr's estimates of the weekly consumption of milk just previous to 1936 by people in six groups classified according to the average income of the group. Each man, woman, and child is counted as a person, and the estimated family income and consumption are divided equally between the members of the family for the purposes of presenting the results. This is a common type of simple diagram and we can see at a glance that (1) the consumption of fresh milk is greater than that of condensed milk in all classes, (2) the consumption of fresh milk increases considerably as income increases, (3) the consumption of condensed milk decreases slightly as income increases, and (4) the changes in total consumption of milk are dominated by the changes in consumption of fresh milk.

An alternative and more austere form of diagram is obtained by using, instead of the rows of milk bottles, long thin rectangles proportional in length to the quantity represented. Some serious-minded statisticians prefer this form of representation and are scornful of 'pictorigrams'. However, there is nothing unsound in their use provided they are correctly done, and I think there is everything to be said for presenting data in as attractive and striking a way as possible, even by using coloured diagrams if they can be afforded. Sometimes, of course, the subject defies acceptable pictorial representation—it would require some ingenuity to represent the death rates of Table 1 or the accident figures of Table 2 by pictures not too macabre for modern taste.

The above data of milk consumption may also be used to show how the method of representation can affect the impression created by the diagram. These same data are represented in Figure 2 by six bottles and six tins proportional in volume to the relative consumptions for the six classes. Figures 1 and 2 present the same set of facts, but the changes in consumption are much less striking in Figure 2 than in Figure 1. It is usually better to adopt a method like that used in Figure 1, where the quantity is represented essentially by a length, since it is difficult

Figure 2.
Relation between Milk Consumption and Income.



to appreciate quantities by areas or volumes, particularly if the latter are inadequately represented on two-dimensional diagrams.

The apparent degree of fluctuation of a quantity can be made to be almost anything we please by choosing the scale of the diagram suitably. This is illustrated by the diagrams in Figures 3 and 4. Figure 3 presents a statistical history of a social phenomenon that for years has been a subject of great public concern, viz. unemployment. The monthly figures of men unemployed at the times of the counts, expressed as percentages of numbers of insured workers, have been taken from the Statistical Abstract, and are represented graphically. The fluctuations seem enormous, but, although they are slightly affected by occasional legislative changes altering the basis of the figures, they substantially represent the actual changes in 'unemployment'. Figure 4 shows

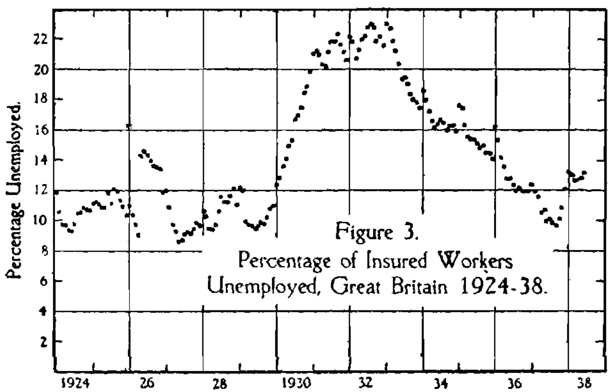


Figure 3.
Percentage of Insured Workers Unemployed, Great Britain 1924-38.

the other side of the same picture and gives the numbers employed at the times of the counts, expressed as percentages of the total insured workers. The fluctuations in Figure 4 seem to be much less than those in Figure 3, but they are actually the same fluctuations, reversed in sign (or turned upside down) and drawn on a different scale. The statistician chooses his scale according to the impression he thinks the figures should convey, and that impression will of course depend on the object for which the figures are being used.

Figures 3 and 4 are typical examples of the well-known *time charts* that are used to depict changes in quantities with time, and are so widely understood that even daily papers with mass circulations use them. From Figure 3 we see the dire effects of the 1931–32 trade depression and the smaller effects of the 1926 General Strike. Figure 4 shows the same effects but

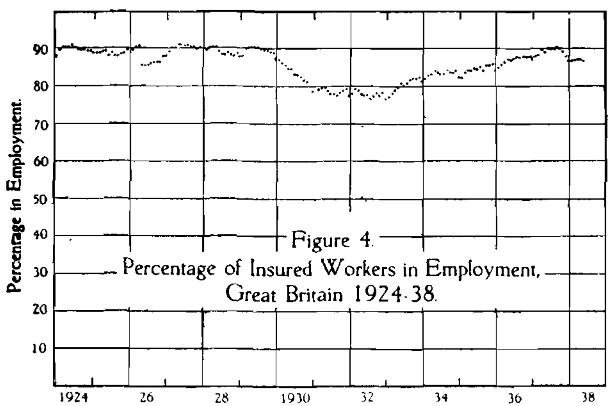


Figure 4.
Percentage of Insured Workers in Employment, Great Britain 1924-38.

puts them in a different perspective, and suggests that these profoundly important changes touched only the fringe of our total industrial effort as measured by the number of people registered as being in work.

The detailed examination of a graph, noting a rise or a fall in level here and there, is a good thing to undertake when there are particular events to which the fluctuations can be related; but more often it is desirable to notice the general character of the fluctuations—to obtain a 'bird's-eye view'. There are several patterns to which time fluctuations may conform, although the combination of two or more patterns may make the resulting form of the graph very complicated. However, it is useful to everybody to be able to recognize the various patterns, and where there are several combined to be able to separate them; so in Figures 5–10 I give a few illustrations and describe the significant features of the changes they represent.[1]

Figure 5 shows an exceptionally steady upward

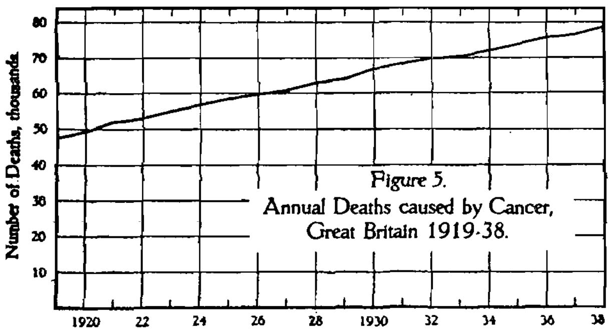
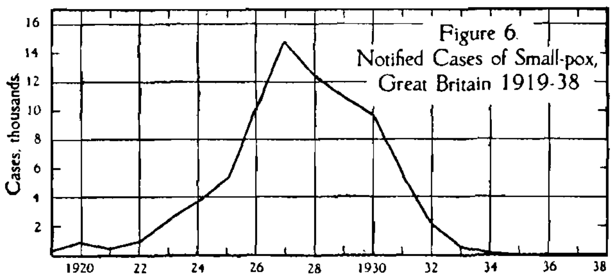


Figure 5.
Annual Deaths caused by Cancer, Great Britain 1919-38.

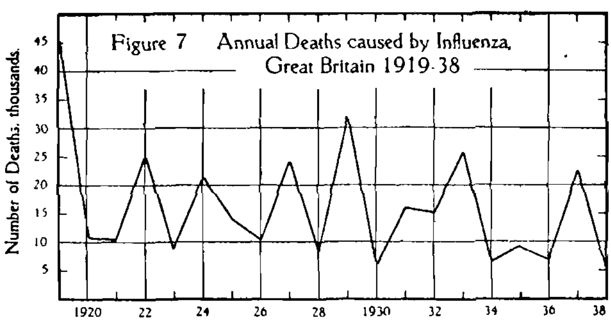
trend in the deaths due to cancer, from about 50,000 in 1919 to nearly 80,000 in 1938. This is partly due to the increasing proportion of older people in the population—older people are more liable to suffer from cancer—and is probably affected also by the increasing accuracy with which cancer is diagnosed. Whether these facts are sufficient to explain the increase in deaths I cannot say.

[1] The data for Figures 5–9 are from the Statistical Abstract, and those for Figure 10 from 'Marriage Frequency and Economic Fluctuations in England and Wales, 1851–1934', by D. V. Glass. This paper is in *Political Arithmetic* (1938).

From Figure 6 we see that the numbers of cases of smallpox rose fairly steadily from a few hundreds per annum in 1919–21 to 10,000–15,000 in 1927–30 (many



readers will remember a smallpox scare in those years) and fell practically to zero in 1935–38. There are also some random fluctuations of negligible importance.



The fluctuations in deaths due to influenza shown in Figure 7 are violent, and at first sight seem to follow no pattern. On closer examination, however, we see a tendency for years with large numbers of deaths to alternate with those with small numbers. Such a pattern, in which the form of the fluctuations repeats itself at regular intervals (two years in this instance), is described as *periodic*. The periodicity 'misses step' in 1921 and 1926, while in 1931 and 1935 the peak is not high, but the general tendency to a periodicity is undoubtedly there. We may sum up the situation by stating that the annual deaths from influenza fluctuated between 5,000 and 45,000, and that usually, but not always, a year with a large number of deaths was

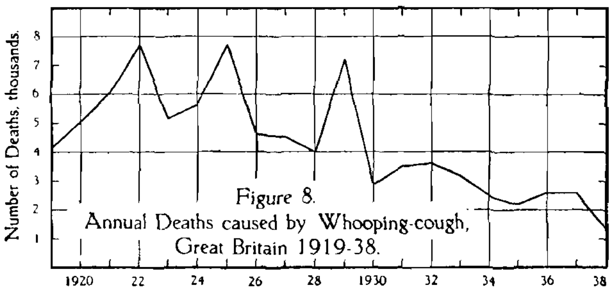


Figure 8.
Annual Deaths caused by Whooping-cough, Great Britain 1919-38.

followed by one with a small number, and vice versa.

The fluctuations in deaths due to whooping-cough (Figure 8) are also large, and are difficult to summarize. I do not know of any factors to which I can relate the variations, and can only describe the main features of the graph as a downward trend from 5,000–6,000 deaths per annum in 1920 and 1921 to about 1,500 in 1938, with occasional 'peak years' in which the deaths rose to 7,000–8,000.

Figure 9 gives a fair idea of the amount of motor traffic on the roads quarter by quarter, and of the numbers of fatal accidents year by year. There is a general upward trend in the numbers of motor licences, the regularity of which is interrupted only by the major effects of the depression years of 1931–32. Superimposed on this trend, however, is a pronounced seasonal pattern. Each year the number of licences rises from a low value in the first quarter (28th February) to a high value in the second (31st May) and a still higher value in the third (31st August);

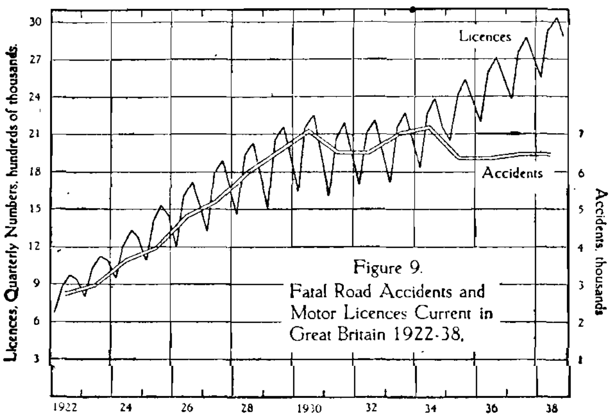


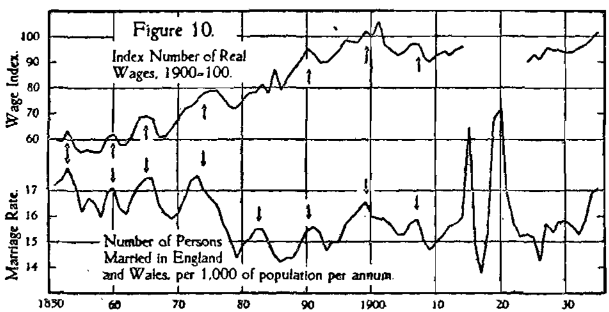
Figure 9.
Fatal Road Accidents and Motor Licences Current in Great Britain 1922-38.

and in the fourth quarter (30th November) there is a slight fall, which is succeeded by a much larger fall in the first quarter of the next year. A seasonal pattern of this kind, which is a special example of a periodic fluctuation, often occurs in data that are given weekly, monthly, or quarterly, although it is not often so well marked as the pattern of Figure 9.

The accidents given in Figure 9 are drawn on such a scale that, if the increases year by year had been in the same proportions as the numbers of licences, the general rise in the curves would have been much the same. Up to 1934 the numbers of accidents and motor licences rose and fell fairly well together. After 1934, however, there was first a drop in the number of accidents, which then remained fairly steady at about 6,400, whereas the number of licences resumed its large annual increase. This relative improvement in the accident figures after 1934 may reasonably be attributed to the conscious efforts, legislative and other, of the nation to reduce accidents; and the result, although not good enough, is encouraging.

The last examples of time series are in Figure 10. The index numbers of 'real wages' plotted there take into account wage rates, the amount of unemployment, contributions to and from social insurance and unemployment funds, and changes in the level of prices. It is difficult to obtain accurate figures on these subjects for as far back as 1850, and how far the index measures the well-being of the workers is not clear, but the results plotted in Figure 10 give rough measure of changes in the prosperity of the workers. The fluctuations in the marriage rates are affected by changes in the proportions of the population that were of marriageable age, as well as in marriage habits. However, it is not our main purpose to speculate on the causes of the fluctuations shown in Figure 10 but merely to note their form.

First, if we pay attention only to the slow, long-period changes, we notice a rise in real wages from the first decade, 1850–60, culminating in a peak soon after 1900, and thereafter a slight fall to a fairly uniform level. Superimposed on this movement is a somewhat irregular wave-like movement, with peaks indicated roughly by arrows. These peaks do not occur at uniform intervals as do those for the influenza data in Figure 7, but the wave-like pattern is well marked. Other economic data show somewhat similar fluctuations, which are described as alternating booms and depressions and constitute the well-known phenomenon of the business or trade cycle. Superimposed on the waves are smaller random fluctuations which produce sporadic peaks and valleys of no great significance (e.g. in 1884–86) and distort the waves.



The *slow movement of the marriage* rate is a *down*ward trend up to about 1880, and thereafter the fluctuations are about a trend which rises very slightly until about 1900. There is a period of wild fluctuation about 1915–20, which was due to conditions caused by the last Great War, and a marked rise in 1934–35 to a value which was maintained until 1938. We see also a wave-like pattern up to the year 1910, with peaks as marked roughly by the arrows, spaced at irregular intervals. I shall later discuss the relation between the two series in Figure 10.

Thus we see that in summarizing and presenting his material the statistician classifies it if necessary, and makes use of suitable tables and diagrams. The picture he gives is impressionistic, and he must do his work skilfully and honestly if he is to avoid creating false impressions. The person who receives the statistical information also has an active part to play. For this he needs to know how to read tables and diagrams, and to understand their meaning; and the more he knows of the general principles on which they are made, the more critically is he able to examine them and the less likely is he to be led astray.

CHAPTER IV

SOME SPECIAL TABLES AND DIAGRAMS OF IMPORTANCE

TABLE 3 gives the weights of fifty apples from the same tree. It is typical of a large class of statistical data in that it refers to a number of things of the same kind, varying in some measurable character, and I propose to show how such figures are dealt with. Usually there are hundreds or even thousands of results in a single collection; there is room here to give only fifty but they will be enough to illustrate the methods.

As they stand, the figures in Table 3 are an almost meaningless jumble, but we can reduce them to order by applying the general methods of classification and summarization mentioned in the last chapter. The weights vary between 68 and 223 grams, and in view

TABLE 3

Weights of Apples, grams

106	107	76	82	109	107	115	93	187	95
223	125	111	92	86	70	126	68	105	130
139	119	115	128	100	84	129	113	204	111
81	131	75	84	104	110	80	118	186	99
136	123	90	115	98	110	78	82	90	107

of this wide range it is obviously unnecessary to distinguish between apples differing by only a few grams. Even two small boys, faced with such a collection and seeing how different the apples can be, might be satisfied that they were being treated nearly enough alike, if one was given an apple weighing

TABLE 4

Frequency Distribution of Weights of Apples

Weight (grams)	*Frequency of Apples*
60– 79	5
80– 99	14
100–119	18
120–139	9
140–159	...
160–179	...
180–199	2
200–219	1
220–239	1
Total .	50

80 grams and the other one weighing say 90 grams. However, we will be content to regard as equivalent, apples differing by up to 20 grams, make a few broad classes covering the whole range of weights, and count the apples in each class. The results are in Table 4. The *sub-ranges*, as they are called, are in the first column, and the numbers or *frequencies* of apples are in the second column of the table. Table 4 is an example of a *frequency distribution*, so called because it shows how the frequencies of apples are distributed between the various classes of weight. It is a summarized form of Table 3, and in obtaining it we have suppressed little or no detail of any importance, even though the classes are very broad.

The two essential elements behind a frequency distribution are the things that are counted, called the *individuals*, and the quantity or quality that is measured and defines the classes, called the *character*. An individual, in the statistical sense, may be a person or a thing; it may be a concrete thing like an apple, or something more abstract like a vote or an experimental observation; and it may be something we ordinarily recognize as a single entity like a man, or complex entity like a family or a business concern. The character of the apples is called *quantitative* because it is described by a numerical measurement, but *qualitative* characters, which are described in words, are also met with. For example, Table 2 (p. 25) gives a frequency distribution in which the individuals are the 5,855 road accidents and the character is the class of road-user causing the accident. However, for quantitative characters the methods of forming a distribution have been standardized, so that a standard interpretation is possible.

In order to discuss what a frequency distribution really means I shall use an example based on more adequate numbers than fifty. The life of an electric lamp is the number of hours it burns at a standard

TABLE 5

Length of Life of Electric Lamps

(Data by E. S. Pearson, *Journal of the Royal Statistical Society*, 96 1933, p. 21)

Life (hours)	*Frequency of Lamps*
0– 200	1
200– 400	3
400– 600	2
600– 800	10
800–1,000	19
1,000–1,200	21
1,200–1,400	23
1,400–1,600	18
1,600–1,800	17
1,800–2,000	10
2,000–2,200	8
2,200–2,400	5
2,400–2,600	5
2,600–2,800	4
2,800–3,000	2
3,000–3,200	1
3,200–3,400	1
Total .	150

voltage, and in Table 5 the results for 150 lamps have been grouped into classes with sub-ranges of 200 hours. For the sake of more vivid representation, this distribution is given as a frequency diagram in Figure 11, where the lamps in each class are piled *in a column proportional in height to the number in* the class. These columns are usually represented by plain rectangles, or alternatively the tops of the columns may be joined by sloping lines; I have used the pictorial form here in order to help readers to understand what a frequency diagram is.

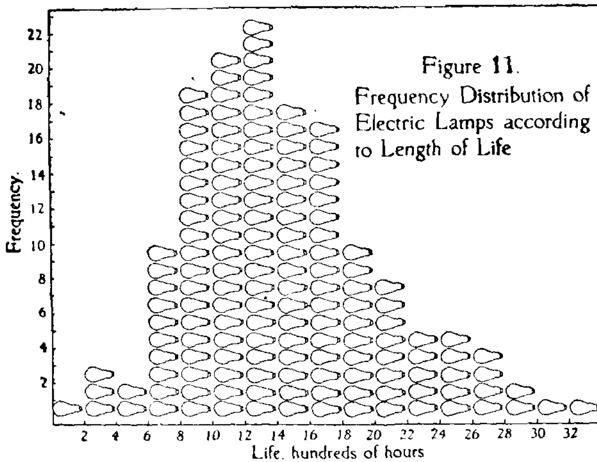


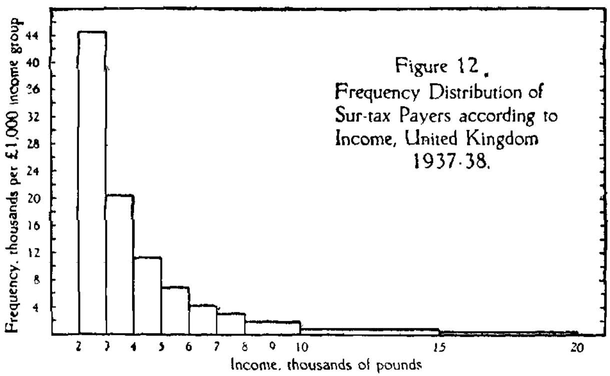
Figure 11.
Frequency Distribution of Electric Lamps according to Length of Life

In considering a frequency diagram, small irregularities in outline such as the minor peak for the 200–400 hour group in Figure 11 are ignored, and notice is taken only of the general shape, which is sometimes shown by a smooth curve drawn through the points of an actual diagram. A diagram of a given general shape means the same thing to statisticians all the world over.

We see that the diagram in Figure 11 is spread from about 0 to 3,400 hours, showing the extent of the variation in life; that there is a peak, showing a tendency for the lamps to be concentrated about a typical value between 1,200 and 1,400 hours; and that there is a reduction in height towards the sides, showing the comparative rarity of lamps approaching the extremes of life.

In order to see one way in which a frequency distribution can arise, let us imagine a shooting target in the form of vertical strips instead of the familiar concentric rings, the centre strip being the 'bull'; and let us consider a marksman shooting at this many times from a rifle. The shots will be peppered over the target. There will be more shots in the central strip than in any other, and as we move from the centre towards the edges each strip in turn will have fewer and fewer shots, until the extreme strips will have very few indeed, or none. If we count the shots in each strip, and draw a frequency diagram, the shape will be like that just described for the lamps, with a peak at the centre and tailing off towards the edges. I do not suggest that all frequency distributions arise in this kind of way, but this illustration often helps readers to see what a frequency diagram means.

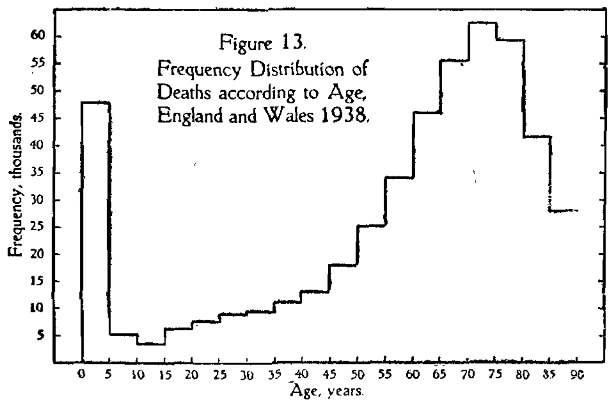
There are two main things to notice about a diagram like Figure 11, after its general shape; these are the position of its peak and its width. The peak for Figure 11 is in the 1,200–1,400 hour group; if we had another batch of lamps with a peak in, say, the 1,800–2,000 hour group, we should probably prefer those lamps as having a longer typical life. The width of the distribution measures the degree of variation about the typical value. For example, the imagined batch of lamps with a typical life of 1,800–2,000 hours might have a distribution ranging from 0 to 4,500 hours, and this batch would be much more variable in life than the original batch. If we had two marksmen shooting at similar targets of the kind just described, a good marksman and a poor one, the frequency diagram for the good one would be narrow with a tall, sharp peak showing little variability in the placing of the shots; that for the poor marksman would be broad and squat, with more of a knoll than



a peak, and showing much variability. This characteristic of variability is a highly important one.

The more or less symmetrical bell-like shape of Figure 11 is the most common in frequency diagrams, but other shapes do exist, which describe other types of variation. For example, the distribution among the people of Great Britain of wealth in most of its forms is such that the vast majority are relatively poor and a very few people are very wealthy; and this distribution is represented by an extremely lop-sided diagram of the kind shown in Figure 12, which shows the distribution of sur-tax payers according to income. The uneven distribution of these incomes is even more pronounced than appears in Figure 12, since if the scale were large enough the diagram would extend with a long thin 'tail' to incomes of over £100,000. Any attempt to represent on the same diagram the distribution of all incomes, including those of the millions of people who do not pay sur-tax, would be defeated by the enormous concentration in the lowest income groups.

The distribution of ages at death in Figure 13, the



mode of representation of which is very slightly different from Figure 12, is of an unusual shape. A relatively large number of those who died were infants, and comparatively few were children over the age of 5 years. From the age of 15 years the numbers increase progressively with increasing age until there is a maximum in the 70–75 age group. Apart from the deaths in infancy, 70–75 years is the typical age at death. The variations shown by Figure 13 are of course the result of variations in (*a*) the death rates, and (*b*) the numbers living in each age group and exposed to the risk of death; and it is because these numbers decrease for the highest age groups that the numbers of deaths in those groups decrease.

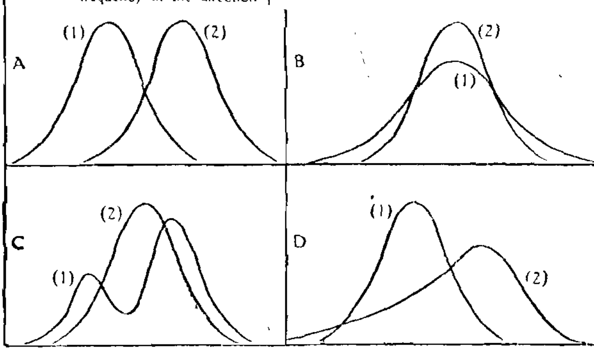
A list of the characters of several hundreds of individuals seems to be an overwhelming and chaotic mass of information, particularly if the items are in some irregular order in which they happen to have been recorded. A frequency distribution is an economical summary of such figures, since it replaces the several hundred readings by a table with some ten or twenty entries, and it reduces the chaos to an order the mind can comprehend. Indeed, the things we need to know about such a body of observations are not really many or complex. We need to know whether there is a well-defined typical value of the character, and if so what the value is; to what extent the individuals vary about that type; and whether the variation extends equally above and below the type. These facts are readily seen from and described by a frequency table or diagram. A roughly drawn frequency curve is often used to give a vivid if approximate description of the nature of variation in a distribution, and with its aid a statistician can sum up a statistical situation, somewhat as a clever cartoonist can, with a few strokes of his pen, convey a whole complex of moods or ideas. In Figure 14 are drawn on the same scale a number of pairs of imaginary frequency curves of some unspecified character, to show the kind of tale they can tell.

If each individual has two characters that have been observed, we may classify the observations according to the two characters, and form a two-way frequency table. If the character is qualitative this is known as a *contingency table*. Table 6 is an example. The individuals are the persons killed in the accidents referred to in Table 2 (p. 25), and the two characters

Figure 14.

Imaginary Frequency Distributions.

The character is represented in the direction → and the frequency in the direction ↑



A. The typical value of (2) is higher than that of (1), but both distributions show the same degree of variation. The overlap of the curves shows that some individuals in (1) have higher values than some in (2).

B. The typical value is approximately the same for (1) as for (2), but (1) shows the greater degree of variation. There are more very high and very low values in (1) than in (2).

C. The distribution (1) shows a mixture of two well-defined types, whereas (2) shows homogeneous variation.

D. The typical value of (2) is higher than that of (1), but the variation in (2) is greater in degree and is asymmetrical in form, so that there are more very low values in (2) than in (1).

classified are (1) the class of road-user the individual was in, and (2) the class of person who caused the accident in which the individual was killed. (There are a few more individuals in Table 6 than in Table 2, since more than one person was killed in some accidents.) We see that 1,064 drivers were killed in accidents caused by drivers, 261 in accidents caused by pedal cyclists, and so on. The entries in the row and column labelled 'Total' form two-frequency

TABLE 6

Numbers of Persons killed in Fatal Road Accidents in Great Britain and Persons to whom the Accidents are attributed, 1936–37

	Accidents attributed to:				
	Drivers other than Pedal Cyclists	*Pedal Cyclists*	*Pedestrians*	*Other Persons*	*Total*
Persons killed:					
Drivers other than pedal cyclists .	1,064	13	17	3	1,097
Pedal cyclists	261	982	25	1	1,269
Pedestrians .	438	51	2,440	1	2,930
Other persons	517	18	2	128	665
Total .	2,280	1,064	2,484	133	5,961

distributions, and the rows and columns in the body of the table subdivide these two distributions, adding to our information.

Table 6 shows clearly two characteristic features of contingency tables. First we notice that the frequencies are not uniformly distributed, most being contained in the cells along the diagonal of the table starting at the top left-hand corner, i.e. in the cells containing the entries 1,064, 982, 2,440 and 128. This shows that most of the people were killed in accidents caused by road-users of the same class as themselves, i.e. there is a strong tendency for the class of person killed and the class of person causing the accident to go together. The second thing we notice is that not all the frequencies are in the diagonal cells: that some of the people who suffered death are not of the same class as those who caused the accident: that there are exceptions to the tendency for the two characters to go together. We may sum up the situation by saying that rough retributive justice seems to have been done between the classes of road-users; justice because the victim tends to be of the same class as the person who caused the accident, and *rough* justice because he is not always of the same class. By the way, before making any attempt to explain the foregoing results, readers should consider two things: (1) the person causing the accident is not necessarily the same individual as the victim even if they are both of the same class, and (2) if the victim of the accident is blamed for causing it he is not there to defend himself.

The general statistical features of Table 6 which I have pointed out are described in statistical language by the word *association*, and we say that the two characters are associated. The word association connotes both the tendency for a connexion between the two characters to show itself, and the deviations from that tendency. In Table 6 the tendency is pronounced and the deviations are not very important, and we say that the association is *strong*.

What would Table 6 look like if there was no association? The answer is: Like Table 7.

The distributions in the 'Total' column and row are the same for Table 7 as for Table 6; they have nothing to do with association. In Table 7, however, the frequencies in each column bear the same relations to each other as those in the 'Total' column, and similarly

TABLE 7

Numbers of Persons killed in Fatal Road Accidents in Great Britain and Persons to whom the Accidents are attributed, 1936–37. No association

	Accidents attributed to:				
	Drivers other than Pedal Cyclists	*Pedal Cyclists*	*Pedes-trians*	*Other Persons*	*Total*
Persons killed:					
Drivers other than pedal cyclists .	420	196	457	24	1,097
Pedal cyclists	486	226	529	28	1,269
Pedestrians .	1,120	523	1,221	66	2,930
Other persons	254	119	277	15	665
Total .	2,280	1,064	2,484	133	5,961

for the rows. For example, 420 is to 486 is to 1,120 is to 254 as 1,097 is to 1,269 is to 2,930 is to 665.

The degree of association may vary from extreme strength, when only one cell in each row and column has a frequency, through the stage shown by Table 6, to the zero association of Table 7. I enlarge below on this idea of strength of association.

In Table 6 the association shows itself by most individuals being in the cells along one diagonal in the table. That is not an essential feature and arises from the particular arrangement adopted of the rows and columns. If the columns, say, are rearranged in any way, the cells with large frequencies will be positioned irregularly in the table, but the association will still be there.

TABLE 8

Numbers of Husbands and Wives of Various Ages residing together on the Night of the Census, 1901, thousands

(Data from *An Introduction to the Theory of Statistics*, by G. U. Yule, 1922)

	Age of Wife								
	15–25	25–35	35–45	45–55	55–65	65–75	75–85	85–	*Total*
Age of Husband									
15–25	193	50	1	..	..	..	..	..	244
25–35	231	1,162	108	4	..	..	..	..	1,505
35–45	12	408	977	92	4	..	..	..	1,493
45–55	1	36	320	652	66	3	..	..	1,078
55–65	..	5	37	211	358	34	1	..	646
65–75	..	1	6	24	105	133	10	..	279
75–85	..	..	1	4	10	30	22	1	68
85–	..	..	..	..	..	2	2	..	4
Total .	437	1,662	1,450	987	543	202	35	1	5,317

When the two characters are quantitative the contingency table takes a standardized form, and the special name of *correlation table* is given to it. Tables 8 and 9 are examples.

The individuals of Table 8 are married couples living together on the census night of 1901, and the characters are the ages of husband and wife. There is a pronounced association which is shown by the tendency for all the frequencies to occur in cells about one diagonal, and this gives the whole table a characteristic appearance.

Association between two quantitative characters is called *correlation*, and we say that the two characters are *correlated*. The strong correlation between age of husband and wife is in accordance with our general experience, for although men sometimes have wives

TABLE 9

Numbers of Husbands and Wives who died at Various Ages. Data from Gravestones in the Yorkshire Dales

(*Biometrika*, 2, 1903, p. 481)

	Age of Wife									*Total*
	15–25	25–35	35–45	45–55	55–65	65–75	75–85	85–95	95–	
Age of Husband										
25–35	2	4	4	5	..	5	2	..	..	22
35–45	..	7	5	4	11	12	7	1	1	48
45–55	..	4	9	7	11	17	28	6	..	82
55–65	..	6	13	17	33	35	32	9	1	146
65–75	1	11	23	18	45	74	64	19	3	258
75–85	3	6	14	17	27	71	81	28	1	248
85–95	..	1	3	7	9	13	21	12	..	66
95–	..	..	..	1	2	..	2	..	1	6
Total	6	39	71	76	138	227	237	75	7	876

much younger than themselves and, less often, wives much older than themselves, age in the husband is usually associated with age in the wife.

If, however, we consider the ages at which husbands and wives die, the correlation largely disappears. Whether the wife dies when young or old makes little difference to the age at which the husband dies. This lack of correlation is shown by the appearance of Table 9, which refers to married couples whose ages at death were recorded on the gravestones of country churchyards in the Yorkshire Dales. Table 9 looks quite different from Table 8 in that the frequencies do not tend to be concentrated about a diagonal, although statistical analysis shows that there is a slight tendency for this to occur and that there is consequently a very weak correlation.

When the number of individuals is small, correlations may be shown to the eye by a *correlation diagram*,

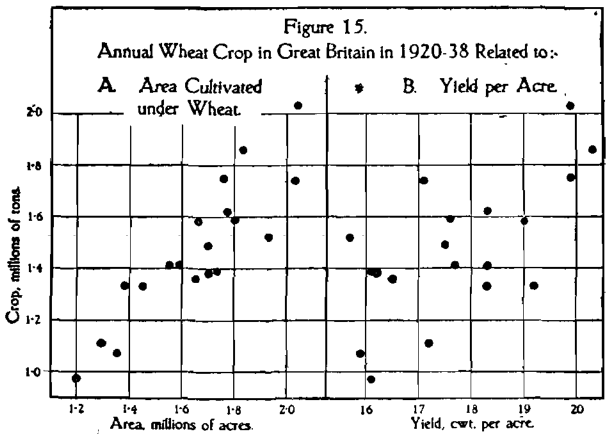


Figure 15.
Annual Wheat Crop in Great Britain in 1920-38 Related to:-

as illustrated in Figures 15A and B. The individuals are the nineteen years from 1920 to 1938, and for Figure 15A the characters are the total wheat crop and the area cultivated under wheat each year. In the diagram, the position of each dot represents for a year the crop and the acreage. In Figure 15B the crop is similarly plotted against the yield per acre. Both diagrams show a correlation by the fact that the points are clustered about a diagonal, but the correlation is stronger in Figure 15A because the clustering is closer and the scatter of the points is less. The concentration of the points in Figure 15A gives it, in essentials, an appearance not unlike that of Table 6.

This conception of correlation was introduced by Galton late in the nineteenth century, and it is now one of the most important and useful ideas we have. Correlation expresses the general idea of a relationship between two quantitative characters and also of departure from that relationship—of a relationship which is apparent and well defined when there is a number of observations, but which describes only approximately the connexion between the two characters for any one individual.

When the two characters tend to increase together, the correlation is said to be *positive*, and when one tends to increase as the other decreases, the correlation is *negative*. In a table or diagram a positive correlation gives the same appearance of clustering as a negative one, but the figures or points tend to follow the other diagonal. The correlations in Table 6 and Figures 15A and B are all positive, and it is perhaps unfortunate that the rules for arranging tables and diagrams should be such that the table shows a 'downhill' trend, moving from left to right, and the diagram an 'up-hill' trend for positive correlation.

We may interpret correlation by imagining that we have taken a man at random, and that we wish to guess or 'predict' his age. If we know nothing about him except that he is one of those recorded in Table 8 we can only say that he is somewhere between 15 and about 90 and that he is most likely to be near the typical age-group of 25–35. If, however, we are told that his wife is between 25 and 35, we see from Table 8 that he will be between 15 and 75. The effect of the correlation is to reduce the range of uncertainty of our prediction. In Table 9, on the other hand, the range of uncertainty of prediction is from 25 to something over 95, if we know only that the man's age is recorded on a gravestone in the Yorkshire Dales; and if we also know that his wife died between 65 and 75 years of age, say, the range of uncertainty is only very slightly reduced. The stronger the correlation, the greater is the accuracy with which a knowledge of one character enables us to predict the value of the other, as compared with the accuracy of the random guess.

Figures 15A and B suggest another interpretation of correlation which is only sometimes valid. It is common sense that the crop will tend to increase with the area cultivated and with the yield per acre; these two factors may be regarded as 'causes' of the crop variation, and the correlation a visible demonstration of their operation. This leads us to an interpretation of the scatter of the points. Had the cultivated area been the only causal factor that varied between 1920 and 1938 all the points in Figure 15A would have been accurately *on* a line sloping diagonally. The scatter is due to the operation of some additional cause or causes, which our common sense tells us can, in this instance, only be those affecting the yield per acre. Similarly, the trend of points in Figure 15B is due to the causes associated with the yield per acre, and the scatter to the disturbing causes associated with the area cultivated. If the correlation is high, as in Figure 15A, the cause accounted for is relatively important and the disturbing causes are relatively unimportant. A lower degree of correlation as shown in Figure 15B may be interpreted to mean that the causes of which account is taken (those associated with yield per acre) are less important relative to the unaccounted causes. The results of Figures 15A and B are incidentally of interest since the causes affecting the changes in cultivated area are mostly economic and political, those affecting the yield per acre are meteorological and technical. In the years 1920–38, although both sets of causes produced variations in the total crop of wheat, variations due to economic and political causes were more important than those due to technical causes and the weather.

Thus we see that when any cause affects the variations in a character, the effect shows itself as a correlation. But the existence of a correlation does not prove the existence of a causal relationship. Two characters can be correlated because they are both affected by a third group of causes, and sometimes they may simply happen to be correlated. Finally, I emphasize the fact that a correlation, like all statistical results, merely describes the relations within a given set of data, referring to a particular set of conditions and taken at a particular time. It may or may not be possible to generalize from such results.

Correlation may show itself in most complicated ways when the two characters are in a time series. Then, the best representation is in a time chart like Figure 10 (p. 37), and it is necessary to consider separately the different features of the time variations. For example, for the years 1850 to about 1890 there is a general upward trend in real wages and a downward trend in the marriage rate, and as far as this trend is concerned there is a *negative* correlation, i.e. an increase in wages goes with a decrease in marriage rate. For the wave-like variations, on the other hand, the peaks of real wages tend more or less to coincide with those of marriage rate and these shorter-term movements show a *positive* correlation, i.e. wages and the marriage rate tend to rise and fall together in the short-term movement. I do not wish to rush in where angels fear to tread and try to explain these results, but they will mean a little more to readers if I suggest tentatively that, as far as the short-term fluctuations are concerned, the wave-like changes in prosperity as measured by the wage index have obvious causal effects on the marriage rate, and that the slow trends in the two series have little connexion with each other, one being largely due to changes in social habits and in the age distribution of the population, and the other largely to changes in industrial efficiency. A correlation between time series that arises from two similar trends is seldom due to a causal relationship; one that results from similar cyclical movements is sometimes due to a causal relationship; and one that results from correlated random movements is often due to a causal relationship. There are, of course, methods of statistical analysis that enable the various kinds of fluctuations and correlations to be separated out and measured more exactly than we can do by a cursory examination of the diagrams. These form a very large subject known as 'The Analysis of Time Series'.

Frequency distributions of single characters give rise to the concept of variation about a type, and tables and diagrams relating pairs of characters to the concepts of association and correlation. These concepts are essential to the statistician but they are also useful to the ordinary citizen, since they help in making sense of figures that come within everyday experience.

CHAPTER V

'EXPRESSING IT IN NUMBERS'

ON a wall of the Biometric Laboratory at University College, London, where much of the present science of statistics has been developed, is written the following motto:

'When you can measure what you are speaking about and express it in numbers, you know something about it, but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind.'—LORD KELVIN.

This motto well expresses the spirit that has inspired statisticians, and much of the work of the pioneers of the subject has been towards developing ways of expressing in numbers measures of statistical concepts which I have so far described only in general terms. I think the motto is an overstatement, since good has come of many qualitative studies—in biology for example—and it omits to state the requirement that the numerical measure should be of a kind that can be brought into a system and related to other quantities, or it is sterile and little better than qualitative knowledge. All the same, the development of numerical measures is a very important step in any science, and in this chapter I describe some of the chief measures in statistics. The use of these measures carries still further the process of summarizing data, bringing into prominence and describing the few features that are of most significance.

The simplest statistical quantities are rates, ratios and percentages. It is not necessary for me to define these here, for we are taught in the arithmetic lesson at school how to calculate them, but I point out one fundamental feature they have in common: they all express the value of one quantity relative to another. A death rate expresses the number of deaths in a locality during a year relative to the number of people living in that locality; and the percentages of unemployed workers are relative to the numbers of insured workers.

One purpose in using such a method of expression is to help people to grasp the meaning of the figures; to bring them home to the *imagination*. In common with most people I have very little occasion to consider the populations of towns and countries, and so when I read that the population of Shanghai is one and a quarter millions, my imagination is unstirred. Some people try to present facts of this kind by some such device as suggesting that 1,250,000 people stood shoulder to shoulder would reach nearly from London to Edinburgh. This does not help me in the least. But I know Liverpool and that it has a population of about 800,000; when I realize that the population of Shanghai is about one and a half times this figure I begin to have some conception of what that is.

One of the jobs of the statistician is to find suitable standards of reference. Note however that the standards should be *suitable*. I remember, a few years ago, hearing the Chancellor of the Exchequer expounding his budget over the wireless and emphasizing its enormous total—some eight hundred million pounds. To those of us who only handle a few hundred pounds of money in a year such figures mean only a tremendous *amount of money*. Had the Chancellor expressed the total as amounting to a rate of about £20 per head of the population, our comprehension would have been a little better, although the impression gained by a man with a family income of £100 per annum would have been different from that gained by a £20,000 a year man. I think the Chancellor would have done best to have expressed the amount of the budget as a ratio of the national income at that time—roughly one-fifth.

A second reason for expressing a quantity as a ratio or percentage of another is that the ratio may contain all the information that matters, actual values of the two quantities being irrelevant details. For example, if we wished to compare the risk of death in two localities, it would be misleading to compare the numbers of deaths, as the populations in the localities might be different, and the sizes of the populations are irrelevant. All the information we would need is contained in the death rates.

Percentages are much used when it is desired to study the relative changes in some quantity with time without considering the absolute amounts of the quantity or the changes. Then the value of the quantity at some given time or *base* is taken as a standard of reference, and the values at other times are expressed as percentages of this. Such percentages are *index numbers*. In studying changes in real wages in Figure 10 (p. 37), for example, we are not interested in the absolute values of the wages, and so we express them as percentages of the wages received in a base year—1900 in this instance. The Ministry of Labour cost-of-living index number is much used as a basis for making wage changes in certain industries. It is calculated from the cost to buy a standard list of goods that enter into the budget of an average working-class family, and the excess of this over the cost in July 1914 is expressed as a percentage of the cost in July 1914. Index numbers are very useful for comparing changes in quantities that differ in kind or magnitude, e.g. for comparing the relative changes in unemployment and exports. It could profitably have been used in Figure 10 had I expressed all the marriage rates as percentages of the rate for 1900.

Some rates or ratios devised for special purposes, particularly those used in population and vital statistics, are very complicated. For example, in studying population trends, a 'net reproduction rate' is calculated from (*a*) the numbers of girls born in a given period, (*b*) the numbers and ages of the mothers, and (*c*) the proportions of the girls that will live to the various ages when they may themselves become mothers. This ratio is so calculated that if it is 1·0 the population is just maintaining its supply of producers of children—i.e. potential mothers. For England and Wales in 1934–36, the net reproduction rate was 0·76, so that there were born only three-quarters of the girls necessary to maintain the population, assuming that fertility and death rates continue unchanged.

Usually, the investigator has little difficulty in devising a reasonably suitable rate, ratio or percentage for his purpose. Thus, the most important cause of the great increase in fatal road accidents between 1922 and 1938 shown in Figure 9 (p. 35) is very probably the increase in the number of motors on the road—in the number of lethal instruments abroad. Indeed, one would expect that if everything else remained constant, accidents would increase proportionately with the number of motors, so that variations in the ratio of the number of accidents to number of motors would indicate the effects of variations in the other factors. However, we do not know the number of motors on the road, but the average number of licences current each year provides a good, if rough, measure, provided we may assume that the annual mileage of the average car did not change

TABLE 10

Number of Fatal Road Accidents per 1,000 Current Motor Licences in Great Britain

Year	*Accidents*	*Year*	*Accidents*
1922	3·2	1931	3·3
1923	2·9	1932	3·2
1924	3·0	1933	3·4
1925	2·9	1934	3·3
1926	3·2	1935	2·7
1927	3·1	1936	2·5
1928	3·3	1937	2·4
1929	3·4	1938	2·2
1930	3·5		

much with time. Thus, a rate that may reasonably be regarded as indicating the effect of factors other than the number of motors on the road is the number of fatal accidents per 1,000 current motor licences. This rate, calculated for the data of Figure 9, is given in Table 10, from which it may be seen that the other factors only became important in effecting the improvement after 1934. Factors that could have influenced the accident rate, but did not change enough to do so appreciably at least up to 1934, include the character of the motors, the numbers of pedestrians and other road-users exposed to the risk of accidents, the skill and care of drivers, and the state of the roads.

Sometimes, however, it passes the wit of man to devise a suitable rate or ratio. In 1938 there were 253 fatal road accidents in Lancashire and 88 in Derbyshire. Does this mean that Lancashire drivers are worse than those in Derbyshire? Not necessarily. The two counties differ in population, in numbers of motors on the roads, and in the length and character of the roads (i.e. in ratio of rural to urban mileage); and I do not know how to devise a rate that will properly take account of these factors and measure the relative standards of driving in the two counties.

A ratio or a percentage is not always the best means of comparing two quantities; a simple difference is sometimes better. A difference in aeroplane speed of say 25 miles per hour is $12\frac{1}{2}$ per cent of 200 m.p.h. and only about 8 per cent of 300 m.p.h., but such a difference is equally important to a bomber trying to outstrip a fighter at both speeds.

Rates, ratios and percentages are rather 'tricky' quantities to deal with, and the unwary sometimes go astray in using them. Most of the errors are due to a neglect of the fact that these quantities are made up of a numerator and a denominator. For example, the percentage of all insured workers that are unemployed is:—

$$\frac{\text{number of insured workers unemployed}}{\text{total number of insured workers}} \times 100.$$

We call this, shortly, the 'percentage unemployment' and so are apt to overlook the denominator, which does not appear in the short title of the quantity. I give a few examples to illustrate this point.

The error of thinking that a large percentage cha in a quantity necessarily means a large actual cha is less commonly made than it was, and when Minister for the Production of This or That bids rejoice because the output of this or that has increas by five hundred per cent, even those of us who a not statisticians sceptically ask 'Five hundred per ce of how much?'

Changes in a percentage may be due to changes i the numerator, in the denominator, or in both; or th changes in the numerator and denominator ma compensate for each other to keep the percentag unchanged. Thus, the percentage of the insurec workers in employment in 1927 was about the same as in 1937, viz. about 90 per cent; but the estimated numbers of insured workers in employment increased from about nine and three-quarter millions in 1927 to rather more than eleven millions in 1937.

The following example is given by Dr. A. Bradford Hill. The data of Table 11 were used by someone to show the bad effect on the infantile death rate in rural districts of the lack of facilities for dealing with confinements. To this lack was attributed the higher percentage deaths under one month in the rural areas given in the last column of Table 11. In fact, as Dr. Hill points out, the death rate under one month of age is lower in rural than in urban areas, and the relatively unfavourable percentage for the rural areas is due to the relatively favourable death rate for all ages under one year, i.e. to the reduction in the denominator of the percentage. A lack of facilities in rural areas may account for some infantile deaths, but the figures of Table 11 do not show such an effect.

Before leaving the subject of percentages protest against the grandiloquent misuse of the term percentage when a simple ratio would be better.

When a ratio is an awkward fraction, it is convenient and legitimate to multiply it by one hundred and

TABLE 11

Deaths per 1,000 Live Births in Urban and Rural Areas, at Various Ages

	Deaths under One Month	*Deaths under One Year*	*Deaths under One Month as Percentage of Deaths under One Year*
Urban .	29·67	95·37	31
Rural .	23·77	58·66	40

convert it to a percentage; 'seven per cent' is a better phrase than 'seven hundredths.' When, however, the ratio is a multiple of unity, it is pretentious to express it as a percentage; it sounds very grand to talk of an increase of five hundred per cent, but it is better English to talk of an increase to six times the original value. Worst of all is a mixture of methods of expression. The mixture is rich in the following extract from a letter to a newspaper, even if we allow that the lack of intelligibility of the last sentence is due to the omission of a £ sign before the last figure:

'As far as British cycles are concerned, the best illustration is that of comparison with Germany, whose exports have fallen in the last few years to under 200,000, whereas during the same period British exports have increased to approaching 400,000, and in the recent trade depression while Germany has lost

The error of thinking that a large percentage change in a quantity necessarily means a large actual change is less commonly made than it was, and when the Minister for the Production of This or That bids us rejoice because the output of this or that has increased by five hundred per cent, even those of us who are not statisticians sceptically ask 'Five hundred per cent of how much?'

Changes in a percentage may be due to changes in the numerator, in the denominator, or in both; or the changes in the numerator and denominator may compensate for each other to keep the percentage unchanged. Thus, the percentage of the insured workers in employment in 1927 was about the same as in 1937, viz. about 90 per cent; but the estimated numbers of insured workers in employment increased from about nine and three-quarter millions in 1927 to rather more than eleven millions in 1937.

The following example is given by Dr. A. Bradford Hill. The data of Table 11 were used by someone to show the bad effect on the infantile death rate in rural districts of the lack of facilities for dealing with confinements. To this lack was attributed the higher percentage deaths under one month in the rural areas given in the last column of Table 11. In fact, as Dr. Hill points out, the death rate under one month of age is lower in rural than in urban areas, and the relatively unfavourable percentage for the rural areas is due to the relatively favourable death rate for all ages under one year, i.e. to the reduction in the denominator of the percentage. A lack of facilities in rural areas may account for some infantile deaths, but the figures of Table 11 do not show such an effect.

Before leaving the subject of percentages, I wish to protest against the grandiloquent misuse of the term percentage when a simple ratio would be better.

When a ratio is an awkward fraction, it is convenient and legitimate to multiply it by one hundred and

TABLE 11

Deaths per 1,000 Live Births in Urban and Rural Areas, at Various Ages

	Deaths under One Month	*Deaths under One Year*	*Deaths under One Month as Percentage of Deaths under One Year*
Urban .	29·67	95·37	31
Rural .	23·77	58·66	40

convert it to a percentage; 'seven per cent' is a better phrase than 'seven hundredths.' When, however, the ratio is a multiple of unity, it is pretentious to express it as a percentage; it sounds very grand to talk of an increase of five hundred per cent, but it is better English to talk of an increase to six times the original value. Worst of all is a mixture of methods of expression. The mixture is rich in the following extract from a letter to a newspaper, even if we allow that the lack of intelligibility of the last sentence is due to the omission of a £ sign before the last figure:

'As far as British cycles are concerned, the best illustration is that of comparison with Germany, whose exports have fallen in the last few years to under 200,000, whereas during the same period British exports have increased to approaching 400,000, and in the recent trade depression while Germany has lost nearly 60 per cent of her exports Great Britain has lost but 36 per cent. . . . So, too, the importance of reciprocal arrangements in the Dominions may best be emphasised by the fact that, whereas in 1929 our exports to the self-governing Dominions and India amounted to £1,461,073, in 1931 the total is estimated not to exceed 400,000.'

When we have a quantity that varies from place to place or from time to time, and we wish to obtain an idea of what the *Concise Oxford Dictionary* calls 'the generally prevailing degree or amount', we calculate an average. The form in which it is usually calculated is known precisely as the *arithmetic mean*, although it is more often referred to in ordinary language as the average. It is the sum of the individual values divided by the number of individuals. There are other averages, but they need not concern us here.

The notion of an average carries with it, by implication, the notion of variation; for we do not average an invariable quantity: we do not ordinarily talk of the average length of a day. When we calculate an average, however, we choose to ignore the variation and focus attention on the 'generally prevailing' value. This means a very big step in the process of statistical summarizing, substituting for the several individual values the one. Sometimes, however, people forget the variation they have ignored, and are misled by taking account of the average alone; and it is because of the prevalence of this error that statisticians are at great pains to stress the inadequacy of this constant. The average has its limitations, but provided they are recognized, there is no single statistical quantity more valuable than the average. When we read that the nation spent an average of 9s. per head per week on food in 1934 and that the average income was about 30s. per head per week, we have conveyed two pieces of information that are striking and useful, even though they tell us nothing of the large variations from one person to another in the expenditure on food and income. The average life of the 150 electric lamps mentioned in Table 5 (p. 41) is 1,452 hours, and if we use such lamps in the home, where we can let each one burn out before replacing it, that average means something. Such lamps at 2s. each are as valuable (as regards life) as lamps at 1s. each that have an average life of 726 hours; and this statement is true irrespective of the variation from lamp to lamp.

When a frequency distribution is, like that of the electric lamps, more or less symmetrical with a peak towards the centre of the range of variation, the average is an important descriptive constant, for it is near the typical value, and the variation is more or less the same above and below it.

Many quantities are in fact averages, although they do not always appear to be so. Thus, in England and Wales in 1938, there were 478,829 deaths in a population of 41,215,000, giving a death rate of 11·6 per 1,000. But the population is made up of people of all ages, following all sorts of occupations and living in many localities, and for every sub-division of the population there is a separate death rate. The crude death rate is an average of all these.

Averaging is very useful in making index numbers. I have already described (p. 60) how, in order to measure changes in a quantity from time to time, the successive values are expressed as percentages of the value at some base period, to form index numbers.

Sometimes, however, the quantity cannot be defined by a single measure, but is (mathematically at least) a somewhat vague and nebulous idea, like the 'price level'. The prices of the things we buy vary from time to time, but in different ways. Some prices rise and fall together, but to different degrees; some prices rise while others fall. For example, between 1920 and 1938 articles like motor-cars had a pronounced downward trend in price owing to changes in the methods of manufacture, whereas the price of coal did not change nearly as much. A good cotton crop may result in a low price for cotton in the same year that a poor harvest results in a rise in wheat prices. Behind all these various movements, however, economists see a movement in the general price level due to common factors that affect the prices of most, if not all, goods in somewhat the same way—factors such as war and money policy. One way of measuring this movement is to calculate index numbers for the separate commodities and then to average them in one way or another. The changes in the index numbers for any one commodity are due to the combined effects of the changes in the general level of prices and the special changes for the commodity; and in the process of averaging for many commodities the special effects tend to cancel out, leaving as the dominating factor in the combined index the changes in general level.

The Board of Trade, the *Economist* and the *Statist* indexes of wholesale prices in Great Britain are obtained in this way. The Ministry of Labour cost-of-living index, already mentioned, is based on retail prices and is obtained in a slightly different way. Broadly, these indexes show much the same fluctuations, but the fluctuations differ in detail; and the economic statistician who knows about the detail can appreciate the economic significance of the differences, and suggest which index is most suitable for any given purpose. There are also other index numbers measuring changes in such quantities as wages, production, industrial activity, and prices of industrial shares. Indexes may also be used for making international comparisons of quantities like real wages.

Let us now consider the limitations of the average. Professor A. L. Bowley has written: 'Of itself an arithmetical average is more likely to conceal than to disclose important facts; it is of the nature of an abbreviation, and is often an excuse for laziness.' The average does not measure the important facts that arise from the variation. In dealing with human problems such as nutrition, for example, it is as important to consider the individuals at the extremes as the average. It is no consolation to the man who can only spend, say, 4s. per week on food that is not sufficient for health, to know that the average expenditure is 9s. per head per week.

We have seen that for the domestic consumer the average life of electric lamps has significance, but some large consumers such as public lighting authorities do not replace lamps as they burn out; they find it more economical to renew all lamps periodically, whether burnt out or not, and for them variability in life is important. Suppose such an authority decided that it would renew lamps at such intervals that only 4 per cent burn out before renewal (it would be very expensive to renew lamps so frequently that none were burnt out). Then for the lamps in Table 5 (p. 41), the renewals would be made after 600 hours of burning; for 6 lamps (=4 per cent of 150) have lives shorter than this. For this consumer the effective life of each lamp would thus be only 600 hours, and not the average life of 1,452 hours. Had the lamps been less variable, the effective life would have been nearer the average life.

Instances in which variation is of practical importance can be found in all fields—the strength of a chain is the strength of its weakest link, not that of the average link; owing to variations in the strength of his materials and in the load a structure will have to bear, the engineer designs the structure with a 'factor of safety'; the banker keeps a reserve of cash in the till to cope with variations in the demand for money; the authority that supplies water allows for variations from time to time in the rainfall when deciding on the capacity of its reservoirs; the electricity supply authority has to cater for a 'peak' load which, owing to variations in demand, is greater than the average load; and so on.

The average of a frequency distribution like that shown in Figure 11 (p. 42), has the merit that it is near the typical value, but when the distribution is like Figure 12 or 13, the average is not even typical. The average income of the sur-tax payers of Figure 12 is about £5,000; the typical income is right at the lower end of the scale, between £2,000 and £3,000. The average age at death of the people who died in 1938 (Figure 13) is 58 years, but such a statement is a very inadequate description of the distribution with its concentrations of deaths at ages under one year, and in the neighbourhood of 70 to 75 years.

When data are in the form of a time series and averages are taken over a long period of time, they are apt to conceal important changes in trend. For example, Table 12, which was obtained by averaging monthly figures of unemployment, suggests an improvement in the situation, continuing up to the end of 1937; but on referring to Figure 3 (p. 30),

TABLE 12

Percentage of Insured Workers Unemployed. Averages for Six-Monthly Periods

Period	*Percentage Unemployed*
Jan.–June 1936	14·2
July–Dec. 1936	12·1
Jan.–June 1937	11·2
July–Dec. 1937	10·5

it will be seen that the percentage unemployment began to increase well before the end of 1937; the average figures conceal this fact.

Even for studying variations, however, the average can be of great use, for we can divide the whole field of investigation into sections and find separate averages for them. Table 1 (p. 23), for example, gives average death rates for the several age groups, and shows the variation with age; the figures of milk consumption represented in Figures 1 and 2 (pp. 27, 29) are averages for the income groups, and show the variation in consumption with income.

Variation has an important effect on the average itself, if that average is a *weighted* one. The average death rate in England and Wales in 1938 was 11·6 per thousand, but this value is not obtained by adding up the nine values given in the top part of Table 1 for

... rate of 54·5 per thousand. The rate of 11·6 results if, when combining the separate death rates, each is given a *weight* proportional to the number of people living in the age group, and could be calculated arithmetically by multiplying the death rates by the corresponding numbers of people, adding, and dividing by the total number of people. The important point to notice is that such a weighted average depends not only on the quantities averaged, but also on the weights with which they are combined; and a change in weights may result in a change of average quite different from that which would result from a change in the quantities alone. I illustrate this from the following data.

The Registrar General divides Wales into two districts: Wales I containing the industrialized counties of the south, viz.: Brecknockshire, Carmarthenshire, Glamorganshire, and Monmouthshire; and Wales II containing the remaining counties which are not as a whole so industrialized. The death rates for 1938 were 12·4 for Wales I and 14·1 for Wales II. In Table 13 I have separated the death rates according to age-groups sufficiently finely divided for present purposes, and we see that in each group, the rate is lower in Wales II than in Wales I except for the one group in which the rates are equal—quite the opposite result from that of the crude total death rates. When we look at the age distributions of the populations in Table 13 we see the reason for this apparent discrepancy. Wales II contains relatively fewer of the younger men and women than Wales I and more people over 55. The death rate in Wales II is relatively high, not because that part of the Principality is more unhealthy or dangerous to live in, but because it contains more aged folk who are relatively prone to die in any locality.

To obtain combined death rates not affected by this difference in age distribution of the population, we

TABLE 13

Death Rates and Percentage Numbers Living at Various Ages. 1938

Age	*Death Rate per 1,000*		*Percentage Number Living*	
	Wales I	*Wales II*	*Wales I*	*Wales II*
0–5	16·5	15·4	7·3	6·8
5–15	1·5	1·5	17·6	14·9
15–35	3·3	3·1	32·2	30·1
35–55	7·5	7·0	26·6	26·2
55–65	23·3	20·6	9·4	11·3
65–75	60·0	49·1	5·1	7·5
75–	150·3	125·6	1·8	3·2
Total .	...	...	100·0	100·0

must obtain weighted averages, using as weights the same population distribution for the two districts. Let us use the distribution of Wales I to give the weights. Then the weighted mean death rate for Wales II corrected to the age distribution of the population of Wales I is $(15{\cdot}4 \times 7{\cdot}3 + 1{\cdot}5 \times 17{\cdot}6 + 3{\cdot}1 \times 32{\cdot}2$, etc.$) \div 100{\cdot}0 = 10{\cdot}9$. This is substantially lower than the death rate of 12·4 for Wales I. This process of correcting a death rate to a standard age distribution is called *standardizing* the death rate. In official vital statistics in Great Britain death rates are also standardised for the sex composition of the population, and the standard population used is one having the same age and sex distribution as the population of the whole of England and Wales in 1901.

After the average of a series of figures, the next thing we usually consider is the amount of variation, ignoring for the time its form. One measure is the *range*, which is the difference between the lowest and highest values in the series, and I use it later in this book because it is easy for the beginner to appreciate. It is not favoured by statisticians, however, except in limited circumstances, partly because it uses only the extreme values in the series and is unaffected by the spread of the values in between. The variation as a whole may be disclosed by measuring the individual values as differences from their average, and may be summarized by averaging these differences, thus obtaining a quantity known as the *mean deviation*. Another measure, called the *variance*, is obtained by squaring and averaging these differences, and the square root of the variance is the *standard deviation*. There are also other, less important measures of variation. The reason for having so many measures and for preferring one to another need not concern us here. It is sufficient to note that the measures are roughly equivalent—for example, they would all show a lower value for distribution (2) in Figure 14B (p. 47) than for distribution (1).

Measures of variation suffer from the same general limitation as the average in that they ignore something—the form of the variation; but since it is not often that we need to compare distributions or series of data for which this form differs much, the limitation has less effect on the usefulness of the measures of degree of variation than on the usefulness of averages. Indeed, the averages and the measures of variation together cover most of the needs of the practical statistician, but their interpretation and use in combination require a good knowledge of statistical theory.

A development of great importance in applied statistics has taken place during the past decade or so, and results from a recognition of the fact that variation is a composite quantity, resulting from the combined effects of a multitude of factors. The combined variation can be broken down into parts associated with groups of these factors and the relative importance of these groups as sources of variation thus be measured. This process, which belongs to a fairly advanced stage of statistics, is parallel to that of breaking down an average death rate, say, into sub-averages for several age groups (Table 1, p. 23).

There are quantities for measuring the form of variation, and formulae including these have been devised for describing the general character of almost all the shapes of frequency distribution that are met with. It is perhaps a consequence of the uniformity of nature and a sign of the achievement of statisticians in condensing and summarizing data that with four constants, including the average and standard deviation, all the essential characteristics of most frequency distributions can be described: that these four constants can contain all the essential data from observations on hundreds of individuals. The measures of form of variation are difficult to interpret in practical work, however, and they are mostly of value in the development of the statistical theory on which practical statistical methods are based.

There are measures of association and correlation, but they require experience and a knowledge of statistical theory for their full appreciation. The most important is the *correlation coefficient* which measures the degree of correlation. If there is no correlation the coefficient is zero; and if there is a perfect correspondence such that changes in one character are exactly related to changes in another, the coefficient is plus or minus 1, the sign merely denoting whether the correlation is positive or negative. A value between 0 and plus or minus 1 describes a degree of correlation between these two extremes; the higher the value of the numeral in the coefficient, the greater is the degree of correlation. Measures of association and correlation are important in practical statistical analysis and with their aid conclusions can be reached from statistical data, that would otherwise be missed.

Rates, ratios and percentages; index numbers; averages, weighted and unweighted; measures of variation; measures of association and correlation; these are among the most important tools of the statistician. Each of these describes some important feature of the data, each leaves much undescribed; each has its uses and its limitations. These quantities should be used carefully, as they are so easy to misuse, and it is perhaps advisable to leave their use mostly to the expert. But anyone may need to understand information expressed by them, and it is well that everyone should know at least something of their meaning.

CHAPTER VI

SAMPLING

THE practice of taking a small part of a large bulk to represent the whole is fairly generally understood and widely used. The housewife will 'sample' a piece of cheese at the shop before making a purchase; and a cotton spinner will buy a bale of cotton, having seen only a small sample of it. The sample is also a very important tool of the statistician.

There are two general reasons for working with samples instead of the bulk. (1) Some appraisals of the thing in question involve destructive tests, and there is no point in appraising it if the whole is destroyed in the process; the housewife cannot eat her cheese and have it. (2) It is very much more economical to investigate a sample than the whole bulk. In social and economic work, for example, it is usually prohibitively expensive to investigate the whole field of inquiry in any detail. Even the population census, which has behind it the financial and coercive resources of the state, is made only at infrequent intervals (the Minister of Health refused to hold a quinquennial Census in 1936 partly on the grounds of expense), and the questions asked are few and comparatively simple. If a sample inquiry is made, on the other hand, it is feasible to employ experienced field workers who can collect information that is comparatively detailed and elaborate, and can ensure that the records are reasonably accurate.

An important example of a sample inquiry is the Ministry of Labour's 1937–38 investigation of working-class family budgets. It was desired, largely for the purpose of constructing a new cost-of-living index number, to know how, on the *average*, working-class families spend their wages. It is inconceivable that all such families in the country could, or would, supply the necessary detailed accounts of their expenditure, so a sample of about 9,000 families was selected, and these were induced to keep very full accounts for four chosen weeks in 1937 and 1938. By working on this scale it was possible for investigators to visit the families and help them to keep their accounts in a way that was more uniform and much more accurate than would otherwise have been possible. As a result, the average weekly expenditure was obtained on about 90 separate items of food, clothing, fuel and light, and other items such as soap and cigarettes, holidays and hairdressing, doctors and dentists, and so on. A more unusual example of the use of sampling is given by Sidney and Beatrice Webb who, during their investigation of English local government, examined all the local Acts in a few selected years as a sample of the thousands of acts passed between 1689 and 1834.

Unfortunately, however, the method of inquiry by sample is somewhat mistrusted, sometimes honestly and sometimes, I suspect, because a sample has in some instance given a result the sceptic does not like. Yet a sample *may* give reliable results. For example, an earthquake disaster in 1923 interrupted the tabulation of the results of the Japanese census of 1920 and interim figures were given based on a sample containing one family in every thousand. These results agreed well with those given later when the regular tabulations were completed. Nevertheless it must be agreed that samples do not represent the bulk exactly, and that they may sometimes be much in error.

Mr. Seebohm Rowntree's 1941 social survey of York was made by visiting all households under investigation, but in his book *Poverty and Progress*, Mr. Rowntree also gives for comparison the results obtained from samples taken from the full returns. He makes no comment on these comparisons (the differences are substantially what a statistician would expect), but a journalist commented on the figures as follows: 'Broadly speaking, they suggest that "sample results" are usually within 15 per cent of the truth either way.' This statement is too broad to have a precise meaning, but in any event it is not the kind of generalization a statistician would make, for he knows that the error in a sample result depends on the size of the sample, on the nature of the bulk being sampled (particularly on the variation within it) and on the way in which the sample is taken. It is my purpose in this chapter to show how this comes about.

In the discussion I shall follow the usual practice of statisticians of referring to the bulk that is being sampled as the *population*. The population in this chapter is to be thought of specially as contrasting with the sample. I shall refer only to populations consisting of recognizably discrete individuals, e.g. men or electric lamps.

The ideal sample is the simple random one in which chance alone decides which of the individuals in the population are chosen. Suppose we wish to obtain a random sample of the people of England and Wales in order to make an estimate of their average height. To do this we may, in principle, take forty-odd million exactly similar cards, one for each person, and write each person's national registration number on the appropriate card. These cards may then be put in a large churn, thoroughly mixed, and (say) one thousand cards be drawn, somewhat in the way the names are drawn for the Irish sweepstake. The thousand people whose numbers are on the cards are a random sample, and we can measure their heights, find the average, and so obtain a figure which is an estimate of the average height for the population.

To investigate the error in the average so estimated we could, again in principle, subsequently measure the heights of all individuals in the population and so obtain the true average. An easier thing to do is to draw a number of samples, each of one thousand, and calculate the several averages. These will vary above and below the true, or population value, and the extent to which they vary gives some idea of the error with which any one sample estimates the true average.

To do such an experiment in fact requires far greater resources than I can command, but there are other experiments that are similar in principle and are easier to do. What we really want to know is how chance works in deciding the choice of the sample, and chance also operates in games of the table, with such things as cards, dice and roulette wheels. In these games, a population does not exist in the sense that the population of England and Wales does, but we may use the concept of a *hypothetical* population. Suppose, for example, we threw a perfectly balanced six-sided die millions of times. We should expect one-sixth of the throws to score aces, one-sixth to score twos, and so on, and the average score would be $\frac{1}{6}(1+2+3+4+5+6)=3{\cdot}5$. These millions of throws are a population, and any thousand of them including the first thousand is a random sample. But the millions of throws need not, in fact, be made; they need only be imagined as a hypothetical population, of which any number of actual throws form a sample.

To illustrate the way in which random sampling errors arise I have made an experiment which I need not describe in exact detail. The experiment is equivalent to that described here, which is not quite so easy to perform but easier to imagine. The imagined apparatus consists of ten packs, each of ten cards, the cards in each pack being numbered respectively 1, 2, 3 . . . 10. The packs are shuffled separately, one card is drawn from each, and the ten numbers on the cards are added to give a score. For example, the numbers might be 2, 4, 2, 10, 2, 5, 9, 2, 9, 8 and the score would then be 53. Then the cards are put back in their packs, the packs are reshuffled, and again ten cards are drawn to give another score. This is repeated, so that a large number of scores results, which are individuals from a hypothetical population consisting of the very large number of scores that could conceivably be obtained. The lowest conceivable score is 10, resulting from ten aces; the highest is 100, resulting from ten tens; and the true average score is 55. Now let us consider the results of the experiment.

It would take too much space to give in full the results of a really extensive experiment, but enough are given in Table 14 to show the kind of thing that happens. The top part of the table gives the first thirty individual scores. Chance has not given a score as high as 100 or as low as 10, as it might have done, and presumably would have done had I continued long enough with the experiment. The first thirty scores vary between 36 and 72, the range being 36. Now, in order to see what happens when we take samples and find the averages, I took 30 samples, each of ten scores. Such samples are far too small for most statistical inquiries (although statisticians sometimes have to be content with small

TABLE 14

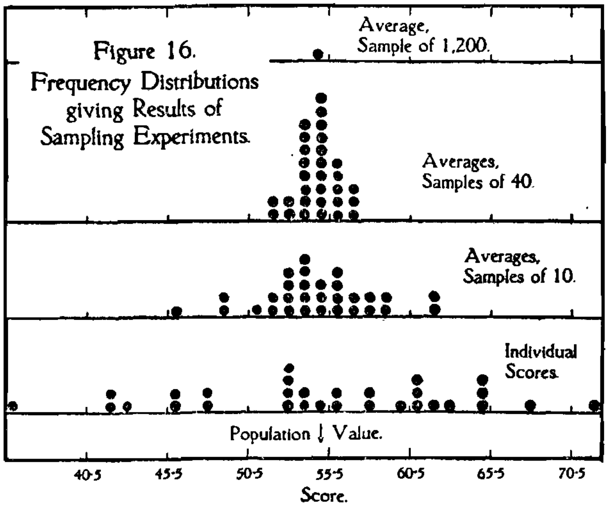
Individual Scores and Average Scores in Samples of Ten and Forty

Individual Scores									
52	46	72	53	36	55	42	56	61	53
56	65	48	54	62	65	48	65	61	60
58	42	58	46	63	61	68	53	54	43
Averages of Samples of Ten									
52·6	58·4	54·6	52·6	48·6	54·0	52·8	50·8	46·0	55·8
53·4	59·4	55·0	56·2	61·6	53·6	54·2	56·8	52·3	54·0
56·7	55·2	56·3	52·3	53·8	57·8	55·9	61·8	58·6	49·2
Averages of Samples of Forty									
54·6	51·6	53·6	56·6	54·3	55·1	57·3	54·4	56·0	55·4
55·3	54·1	55·8	55·4	56·0	53·2	55·1	54·3	54·8	54·2
54·3	57·2	53·2	56·0	54·5	51·5	53·7	56·0	54·8	55·4

samples) but they illustrate the errors of random sampling. The average scores are in the middle section of Table 14. The first average of 52·6 is obtained from the ten individual scores in the top row of the table. The thirty averages vary between 46·0 and 61·8, the range being 15·8, and no average differs from the population value of 55 by more than 9·0. In so far as these thirty samples show the variations we are likely to get in the averages of the millions of samples we could draw, we may say that the biggest error with which the average of any one sample of ten scores estimates the population average is 9·0. When I took larger samples, each of forty scores, I obtained results given in the lowest section of Table 14. They vary between 51·5 and 57·3 with a range of 5·8, and the biggest error with which any one sample of forty scores estimates the population average is $55-51 \cdot 5=3 \cdot 5$. Thus we see that the averages estimated from random samples vary among themselves and differ from the average for the population, but that the biggest error decreases as the size of the sample is increased from ten to forty; and you may take on trust that this tendency would have continued had I extended the experiment to deal with still larger samples. For example, by calculating the average of the thirty averages of samples of forty, we have the average of a single sample of 1,200 scores, which comes to 54·8—very close to the population value of 55.

These results are shown in the frequency distributions of Figure 16 where, instead of a frequency for each sub-range, there are dots, each dot representing an individual score or the average of a sample. Notice how the averages tend to be clustered more closely round the population value as the size of the sample is increased. A frequency distribution of sample averages for any given size of sample is called the *sampling distribution* of the average.

The errors of random sampling, which in an experiment like that just described show themselves as variations between sample means, arise from the variation between the individuals in the original population. Other things being equal, such sampling errors are proportional to the amount of variation in the population. As an extreme example, it is easy to see that had there been no variation between the individual scores and they had all been 55, the means



of all samples of all sizes would have been 55 and there would have been no sampling errors.

When the statistician thinks of the random error of the average of a sample he thinks of a whole collection of possible values of error, any one of which the given sample may have: of the sampling distribution of errors. The actual error of the given sample probably exceeds the smallest of these values; it may easily exceed the intermediate values; and it is unlikely to exceed the very largest values. There is a whole list of probabilities with which the various values of error are likely to be exceeded, and these can be calculated from a quantity called the *standard error*. The standard error is a measure of the variation in the sampling distribution analogous to the standard deviation (p. 74) and for the statistician it sums up the whole distribution of errors. If the standard error of a sample is large, the errors to which that sample is liable are, as a whole, large; if the standard error is small, the likely errors are small. This quantity, carrying with it the idea of errors occurring with various probabilities, should replace the cruder 'biggest error' I introduced in describing the results of the experiment.

It is not usually necessary to do an actual experiment to measure sampling errors, as the mathematical theory of probability enables statisticians to deduce sampling distributions and standard errors theoretically. This method is better because it is less laborious and more exact, giving results as accurate as an experiment involving millions of samples. The results of the theoretical calculations are of the same kind as those given by the experiment, and some instances they have been checked by very large-scale experiments.

I have considered only the sampling errors of the average, but the same principles apply to other statistical quantities such as ratios, and the measures of variation and correlation. The theoretical deduction of sampling distributions of the many statistical quantities in use is a very highly developed branch of mathematical statistics; and sometimes the problems have proved so difficult to solve that statisticians have had to fall back upon actual sampling experiments.

With the ability to calculate errors of sampling, statisticians can make allowances for them when making deductions from sample results. It is a standard procedure to examine the results of a sample to see how far they can be explained by random errors. This is called *testing the significance* of the results, and only such results as cannot reasonably be attributed to errors of random sampling are held to be *statistically significant.*

Before going on to the more practical problems of sampling, I will summarize the ground covered so far. When many samples of the same size are taken from a population of variable individuals, the sample averages show variation which may be described by a sampling distribution and measured by the standard error. A given sample of that size may have any one of the averages in the distribution, and the probability that its error will exceed any stated value can be calculated from the standard error. The standard error of the average is a measure of the errors to which a sample average is liable. For a sample of given size, this standard error increases as the variation between the individuals in the original population increases; for a given population, the standard error becomes smaller as the size of the sample is increased. (For the sake of the mathematically minded it may be stated that the standard error is inversely proportional to the square root of the number in the sample.) Consequently the random errors can be made as small as we please by making the sample large enough, and for a given population it is possible to calculate the size of sample necessary to reduce these errors to any desired value. Similar remarks apply to quantities other than the average.

The tendency for large samples from some population to have averages that vary little amongst themselves and differ but little from the population value is the reality behind the popular conception of the Law of Averages. This law does not operate, as some people think, so that an abnormally high individual score or run of scores is followed by an abnormally low score or run, correcting the average by compensation. In a random series, the scores following an abnormal score or run are quite unaffected by what has gone before; they tend to be nearer the general average than the abnormal scores are, i.e. to be more normal, so that when included in the average they reduce the effect of the abnormal scores. Averaging has more of a swamping than a compensating effect. Thus, if we may regard the days of weather as individuals from a population, the average weather for the population being the general type experienced at a given time of the year and place, the law of averages does not require that a very wet spell shall be followed by a very dry spell. For all I know, there may be a law to that effect, but if so, it is not the law of averages.

If the individuals in a statistical population are well mixed up, no known method of investigation can give more accurate results for a given cost than the method of purely random sampling just described, unless something is known about the individuals to enable some sort of selection to be made. Sometimes, however, a more complex form of random sample called the representative sample gives greater accuracy. Suppose, for example, that in a housing survey we wish to find the average number of rooms per family in some town. Some families at one end of the scale of wealth will live in one room each and at the other end there may be families that have say twelve rooms each; and this variation over a range of eleven rooms per family will give rise to a certain standard error in a simple random sample. Suppose further that we can divide the town into three districts—'poor', 'middle-class', and 'wealthy'—in each of which the total number of families is known, and that the range of variation of rooms per family is from one to seven in the poor, from four to ten in the middle class, and from six to twelve in the wealthy district. Then if we take a random sample from any one district, the district average is estimated with a smaller standard error than that just mentioned, resulting from a range of variation of six rooms per family (i.e. 7 minus 1, 10 minus 4, or 12 minus 6). Further, it can be proved that if a *representative* sample of the same size is taken, in which the proportion of families from each district is the same as in the whole town, the standard error of the average of that sample will be the same as the smaller error resulting from a range of variation of six rooms per family. This is because the proportion of families from each district is left to chance in the simple random sample; in the representative sample it is not, and that source of error is removed.

Random sampling is the basis of the representative sample, however, which is nothing more than a weighted combination of random sub-samples.

Representative sampling is used in the Gallup polls of public opinion, where care is taken to see that the opinions of various classes of people are represented in appropriate proportions instead of leaving it to chance to determine what these proportions shall be.

If it is granted that the ideal random sample can be a reliable instrument of investigation, the questions remain: Can the ideal be attained? Are the actual samples that are used as reliable as random samples?

As a random sample is increased in size, it gives a result that progressively comes closer to the population value, whereas samples taken in some of the ways that are used give results that progressively come closer to some value other than the population value, results that may for some kinds of samples be too high, or for others too low. A sample of this kind is said to be *biased*, and the difference between the value given by a very large sample and the corresponding population value is called an *error of bias*. A biased die, for example, is one for which the fraction of throws showing an ace, say, tends to a value other than one-sixth (the value for the hypothetical population), and the greater the number of throws, the clearer is it that the fraction of actual aces is not one-sixth. Errors of bias are added to the random errors, and since they follow no laws from which they can be calculated, they must be eliminated entirely or reduced so that they become unimportant. This may be difficult to do, and it is often necessary to use very elaborate sampling methods to avoid errors of bias.

It is nearly impossible for anyone to select individuals at random without some randomizing apparatus. If a teacher tries to select a few children from a class, he will tend to choose too many clever ones, or dull ones, or average ones; or if he tries to be random he may select too many clever and dull children and too few intermediate ones. In selecting a sample of houses 'at random', the investigator will be very unlikely to select anything like the right proportions of large and small ones, shabby and smart ones, new and old ones, and so on. Bias almost inevitably will creep in. This is illustrated by the results of large experiments conducted on several thousands of school children in Lanarkshire in 1930 to measure the effect of feeding them with milk, on their growth during the period of the experiment—about six months. At each school the children were divided into two comparable groups; one group received the milk and the other did not, and the effect of the milk was to be measured by comparing the growth rates of the two groups. The results for a number of schools were combined. In an experiment of this kind, the accuracy depends very much on the two groups or samples of children being similar on the average before the feeding with milk begins, i.e. on one being unbiased with respect to the other. To secure this, the children were selected for the two groups either by ballot or by a system based on the alphabetical order of the names. Usually, these are both good ways of making unbiased random samples of the two groups, but the whole thing was spoilt by giving the teachers discretionary powers, where either method gave an undue proportion of well-fed or ill-nourished children, 'to substitute others to obtain a more level selection'. Presumably the substitution was not done on the basis of the actual weights of the children, but was left to the personal judgement of the teachers. The result was that at the start of the experiment, the children in the group that were later fed with milk were smaller than those in the other group, the average difference being an amount that represented three months of growth. It has been suggested that teachers tended, perhaps subconsciously, to allow their natural sympathies to cause them to put into the 'milk' group more of the children who looked as though they needed nourishment. This bias did not ruin the experiment, but unfortunately the interpretation of some of the results was left somewhat a matter of conjecture instead of relative certainty, and there was later a certain amount of controversy about some of the interpretations. The substitutions of the children could have been done without introducing bias had the actual weights been made the basis, and there would have been an improvement on the purely random sampling; but by unwittingly introducing the bias, it seems that the teachers actually made matters worse.

A sampling method that is very liable to give biased results, particularly when testing opinion on controversial matters, is that of accepting voluntary returns. An undue proportion of people with strong views one way or the other are likely to make the returns, and people with moderate views are not so likely to take the trouble to represent them. For this reason, the post-bags of newspapers and Members of Parliament do not give random samples of public opinion.

A spectacular example of a biased sample is provided by the attempt of the American magazine, the *Literary Digest*, to forecast the results of the Presidential election of 1936 by means of a 'straw vote'. Some ten millions of ballot post cards were sent to people whose names were in telephone directories and lists of motor-car owners, and several million cards were returned each recording a vote for one of the candidates. Of those votes, only 40·9 per cent were in favour of President Roosevelt, whereas a few weeks later in the actual election he actually polled 60·7 per cent of the votes. Those from among telephone users and motor-car owners who returned voting cards did not provide a random sample of American public opinion on this question.

Bias does not result only from obviously bad sampling methods; it may arise in more subtle ways when a perfectly satisfactory method is modified slightly, perhaps because practical conditions make this necessary. In some Ministry of Labour samples of the unemployed, a 1 per cent sample was made by marking every hundredth name in the register of claims, which was in alphabetical order. Bias was introduced by not confining the inquiry to the marked names; instead, the first claimant appearing at the Exchange whose name was marked or was among the five names on either side of the marked one, was interviewed to provide the necessary data. Claimants who are in receipt of benefit attend at the exchange several days in a week, whereas those whose claims are disallowed but who are maintaining registration only attend once a week. The effect of this and of the latitude allowed in the choice of persons for interview was that too many claimants in receipt of benefit were included in the sample. It was only when the existence of this bias was realized that some of the results that were apparently inconsistent with other known facts made sense. A similar kind of effect can arise in surveys of households if no one is at home when the investigator first calls at some house chosen to be one of the sample. Such houses are likely to contain small families with few or no young children, since in large households someone is almost certain to be at home to answer the door; and unless the houses with no one at home are re-visited, the sample will be biased in respect of size and character of household.

Although there is no general theory of errors of bias by which the amount of such errors can be calculated in any particular instance, as can be done for random errors, statisticians do not work entirely in the dark.

Sometimes the sample gives, as part of its results, information that is also known accurately from a full census, and the sample is usually regarded as free from bias in all respects if in this one respect it agrees with the census. The soundness of the results of a sample inquiry may sometimes be checked by comparing them with data obtained in other ways, perhaps by other investigators. Where none of these checks are available, it may be necessary to rely on the statistician's general experience of sampling methods in deciding whether the sample in question is a good one. I have given enough examples to show that a good deal is known of the ways in which errors of bias arise, and what must be done to avoid them.

It is implicit in my definition of errors of bias that they cannot be 'drowned' by taking very large samples, in the way that random errors can; a fact that the experience of the *Literary Digest's* straw-vote on the American Presidential election of 1936 amply confirms. From this point of view, a good sample can be arrived at only by employing a good sampling method. I have already mentioned some methods incidentally, and it is only necessary here to give it as a warning that when a statistician advises adherence to an elaborate method with a closeness that may seem to the layman to be 'fussy', that advice had better be followed; failure to do so has been known to lead to biased results.

Altogether, the method of inquiry by sample is difficult and full of pitfalls. But statisticians could not get on without it and experience of its use is both wide and deep, so that in competent hands the method is capable of giving results that are reasonably accurate. Moreover, the inevitable errors in the results can be estimated, and allowance can be made for them in arriving at conclusions.

CHAPTER VII

TAKING ACCOUNT OF CHANCE

CHANCE operates in many fields besides that of random sampling, and many of its effects can be calculated by applying the same general methods as are used to calculate the errors of random sampling. Some of the further applications of those methods will be described in the present chapter.

The effects of chance can be calculated only because they follow certain laws, but these differ in kind from the *exact* laws of subjects like physics. Events that follow exact laws can be described or predicted precisely; but we can only specify probabilities that chance events will occur, or specify limits within which chance variations will probably lie. Newton's laws of motion, for example, are exact because they describe exactly the relations between the motion of bodies and the forces acting upon them; the errors of random sampling follow chance laws because we cannot predict exactly what average a random sample will have; we can only state, as I have suggested on p. 84, the probability that it will lie within certain limits.

I cannot embark upon a full discussion of what we mean by chance, but as a preliminary I shall indicate a few ideas associated with the word. Statisticians attribute to chance, phenomena (events or variations) that are not exactly determined, or do not follow patterns described by known exact laws, or are not the effects of known causes. That is to say, the domain of chance varies with our state of knowledge—or rather of ignorance. Such ignorance may be fundamental because the relevant exact laws or causes are unknowable; it may be non-essential or temporary, and exist because the exact laws do not happen to have been discovered; or the ignorance may be deliberately assumed because the known exact laws and causes are not of such a character that they can profitably be used in the particular inquiry in hand.

An example of ignorance that, according to present-day ideas, is fundamental, is in the Principle of Indeterminacy of modern physics; we do not and cannot know the precise motion of an electron. We do not know what determines the position of a shot on a target, but that ignorance is non-essential and in some degree temporary. The variations in the positions of the shots depend on a host of factors such as variations in the primary aim of the marksman, the steadiness of his hand, the weight, size, and shape of the bullets, the propelling charges, the force of the wind, and so on; but presumably these factors can be investigated and laws be discovered. Indeed, this has happened; and the history of gunnery shows the temporary character of the ignorance. Gunnery is much more of a science and more exact than it was in the days of the Battle of Waterloo, or even during the 1914–18 War; and as knowledge has increased, unpredictable variations in placing of shots have been reduced; but at each stage these variations are regarded as due to chance. Ignorance of causes is assumed by an insurance company in using its past experience of accident claims to establish future premiums for motor-car insurance. The company has considerable knowledge of the circumstances surrounding every accident on which a claim is made, but is unable to make more than limited use of that knowledge, and so treats accidents largely as chance events, except for a few special allowances such as 'no claims bonuses' or extra premiums charged to people with bad accident records.

Usually, events regarded as coming within the domain of chance are those governed by a complicated system of many causes, each of which produces only a small variation; and one frequent characteristic of such events is that small changes in the circumstances surrounding them make a big difference to the results.

Chance as I have described it operates in a very wide field, covering the whole of the unknown; but mathematical calculations can be made and chance laws be propounded only for comparatively simple systems covering a portion of this field. Nevertheless, such calculations have a wide range of usefulness, which the following examples will illustrate.

One use of chance calculations is for deciding which of the fluctuations in a time series are random and which are trends having some significance. As an example of a time series, consider the unemployment data represented in Figure 3 (p. 30). Readers will have no difficulty in recognizing the broad changes, viz. the minor waves in 1925 and towards the end of 1928, the large upward sweep in 1930, the improvement from the end of 1932 to the end of 1937, and the upward movement again in 1938. For the time being we shall omit 1926, the year of the General Strike, as being exceptional. These changes are reasonably attributed to fundamental causes that operate fairly slowly and may be represented by a smooth curve drawn through the actual points of the graph. There are a number of mathematically determined curves that have the property of changing in level in such a slow, regular way, and are of the nature of exact laws or descriptions.

Let us imagine such a curve to be drawn through the points of Figure 3. Then the actual points will be seen to deviate from this curve. They may in some degree follow a seasonal pattern (another exact law), but in Figure 3 that pattern is not very evident, and most of the deviations are irregular. Presumably many of them can be explained in terms of a minor strike in some industry, a political change in some country affecting another industry, an exceptionally hard winter, and so on; but we cannot bring such knowledge into a system, and so we assume ignorance. Hence, having tried all the known kinds of exact laws that are relevant, we consider the deviations to be due to a complex system of chance causes that operate we know not how; and we apply to them the same laws as describe the results of games of chance and sampling experiments. This is the argument for applying the theory of errors of random sampling to testing the statistical significance of fluctuations in time series. For example, there was a sudden and temporary rise in unemployment in the beginning of 1936; is it significant? It actually occurred, and therefore is real, but when we ask the above question we in effect ask: Can the rise be reasonably considered as a random fluctuation arising from that system of causes we have labelled *chance*, or has some unusual event happened? And so we apply the theory of random errors. If this theory had been applied to testing the statistical significance of the sudden rise in unemployment in 1926 it would have shown the operation of something unusual—we know that to be the General Strike.

This kind of application of the theory enables us, in retrospect, to decide whether any particular events with which we try to associate fluctuations have had important effects compared with the system of random fluctuations. When following changes week by week or month by month as they occur it is useful, too, to be able to decide whether the last increase or decrease is large enough to call for action, or whether it is random. At one period, for example, a local newspaper used to publish weekly figures of deaths due to road accidents in a certain town, and the number used to fluctuate about an average of four or five per week. Should we worry if between two particular weeks the number rises from three to six, or rejoice if it falls from five to two? No! Such changes are no greater than any that can be attributed to chance, and do not indicate a real change in conditions. Sometimes the chance coincidence of random fluctuations may give rise to several consecutive small increases or decreases, giving a spurious appearance of a trend. Sampling theory can show when such is the case.

To arrive at results of these kinds, it is necessary to analyse the time series so as to separate the random fluctuations from the secular movements; and additional complications occur if the system of random fluctuations changes. Some would say, for example, that in trade the random fluctuations during a slump and a boom are different. The whole analysis is only approximate, but it is based on ideas that are sufficiently close to reality to give useful results.

The theory of random errors was used for measuring the accuracy of astronomical measurements long before it was applied to statistical samples, and it is used somewhat in measuring the errors of experimental observations in general. When the astronomer measures, say, the position of a star, he finds that in spite of the precision of his apparatus, and the care with which he adjusts it and makes his observations, he does not get the same answer from successive determinations. He repudiates the idea that the position is varying and attributes the variations in his results to unavoidable errors of observation. The question arises: What is the true position? And if it cannot be measured exactly, how accurately can it be estimated? A similar situation arises in the other so-called exact sciences: e.g. in physics and chemistry. Several determinations have been made of the velocity of light, but they do not agree exactly; and a chemist would be very surprised if he got exactly the same result every time he measured an atomic weight.

This interpretation of experimental results as being due to an invariable quantity plus observational or experimental errors is purely a mental conception. The only reality is the set of observations, the characteristics of which can, if desired, be expressed by any statistical constants such as the average, or a measure of variation, or by a frequency distribution. For most experiments, however, it is useful and (within limits) valid to adopt the more common conception.

The errors do not follow any known exact laws, and so the laws of chance are sometimes used to describe them. In applying these laws, the results are regarded as a random sample from a hypothetical population of results, the average of this population being the true value. Then, the average of the sample is an estimate of the true value, and the error in that estimate can be calculated as for any statistical sample. Is this idea valid? On the face of things, it seems as reasonable to imagine the millions of results that would have been obtained had the experiment been repeated millions of times under the same conditions as to imagine the results of millions of tosses of a die. But it is not so certain that the variations between experimental results are entirely of the same kind as those we get when we toss dice.

On this question there are differences of opinion among experimentalists. Some refuse to admit any similarities between experimental and random errors. Others, faced with otherwise intractable results, use the theory of random errors as the only way out. Experimental errors are not, in general, random. There are 'personal' factors, and any one person shows a bias that changes from time to time. I prefer to regard a set of experimental results as a biased sample from a population, the extent of the bias varying from one kind of experiment and method of observation to another, from one experimenter to another, and, for any one experimenter, from time to time. If this view is accepted, experimental errors can be regarded as forming a chance system, but the system is not as simple as that assumed in calculating the errors of random sampling.

In general the bias cannot be estimated and the theory of random errors is therefore not enough. Sometimes, however, one can say that the bias is likely to be small compared with the random errors, and then the theory may give useful, if approximate, results. For example, if, say, five separate chemists were to determine the atomic weight of an element independently, in different times and places, and possibly by different methods, the results would vary because of the effects of random errors and bias. But the separate biases for the five chemists would differ and so would appear as the random errors between the results, the group as a whole would probably exhibit but little bias, and the theory of errors would provide a reasonably close measure of the precision with which the average of the five results estimates the true atomic weight. This might not be so, on the other hand, for the average of, say, twenty consecutive determinations made by one chemist in one laboratory.

Errors of bias are often relatively unimportant when the observed quantity is the difference between two similar quantities. In measuring the distance between two lines in a spectrum, for example, the main error is often due to the uncertainty of setting the cross-hairs of the measuring microscope on the centres of the lines. If there is a bias in doing this, it is likely to be similar for the two lines (provided they are not too dissimilar in width and appearance), and the difference in the two settings will probably be practically unbiased. The theory of errors gave a result that was at least qualitatively right, when applied to Lord Rayleigh's measurements of the density of nitrogen. He made a number of determinations on 'atmospheric' and 'chemical' nitrogen and found a difference in the two averages. Subsequent treatment by the theory of errors has shown that the difference is greater than can be attributed to random variations, and this result is in accordance with a real difference we now know to exist, owing to the presence of the rarer inert gases in 'atmospheric' nitrogen.

Where the bias is completely unknown, I doubt if it is possible to do more than hope that the true value lies somewhere between the highest and lowest of the actual values, and regard the average as an estimate of the true value, that is as good as, but no better than, any other single estimate that could be made from the data. It is, of course, the experimenter's job to reduce bias and random errors to a minimum.

To sum up, the theory of random errors may be usefully applied to some experimental observations, particularly of differences in values, but great caution must be observed on account of bias. Certainly such an application is no substitute for careful experimental control.

Much experimental work, particularly in biological subjects, is now done under conditions, many of which can be well controlled, and the observations can be made accurately; but the material is inherently variable and the results have to be treated statistically. The Lanarkshire experiment made to measure the effect of milk on the growth of children, already mentioned on p. 90, is of this kind. The amount of milk fed can be controlled, children fed and not fed with milk can be kept in the same environment, and the changes in weight can be measured accurately; but it would not do to base conclusions on an experiment on, say, two children. Children vary, and it is necessary to observe a large number and take averages.

The problem of interpreting the results of such experiments is essentially statistical, and it has fallen to the lot of statisticians to study the general questions of arranging experiments with variable material, of drawing conclusions from the results, and of testing them. Under the leadership of Professor R. A. Fisher, who started this work at the Rothamsted Experimental Station (for agriculture), an elaborate technique for doing this has been developed and is very widely used. I propose to give some description of this subject.

There are three main principles to be observed in designing such an experiment; they are replication, randomization, and economy in arrangement.

The necessity for replication has already been stated. The problem first arose chiefly in agricultural field trials made to measure such things as the effects of various fertilizers on wheat yield. It was early seen that different plots treated in the same way gave different yields. Hence, it was not sufficient to have two plots, say, to treat one with a fertilizer, to grow the crops and measure the yields, and to regard the difference as measuring the effect of the fertilizer. The experiment had to be replicated by treating several plots in each way and measuring the difference between the average yields.

Even differences in such averages can be affected by variations between plots, as we can see from the results of the sampling experiment described in the last chapter; and it is desirable to estimate the accuracy of the observed difference. The only known way of doing this is by the theory of random errors. It was found, however, that variations in plot fertility were not random. There was usually a fertility pattern, e.g. a gradient in fertility across the field. In order that the theory of sampling could be applied, an element of randomization was introduced artificially by using some such device as a ballot to decide which plots should receive the various experimental treatments. This is a 'trick of the trade' for making fertility variations into a comparatively simple chance system. A statistician might apply this principle to the above-mentioned experiment of feeding milk to school children by tossing a coin once for each child, giving that child milk if the result is 'heads', say, and no milk if the result is 'tails'.

The pattern in fertility differences between plots in a field was used to increase the accuracy of experimental comparisons. Adjacent plots tend to be more alike than those in different parts of the field, and by comparing the treatments on adjacent plots the random variations affecting the comparison were reduced, with an increase in accuracy. The other way of increasing accuracy is to increase the number of plots, and hence the expense of the experiments; the arrangement using adjacent plots is therefore more economical. In the same way, had it been possible in the Lanarkshire milk experiment to use identical twins, giving milk to one of each pair, far fewer children would have given the same accuracy as thousands chosen at random. This kind of arrangement can be made to satisfy the condition of randomness sufficiently for the application of the theory of random errors in an appropriate form.

The above are the elementary principles of the modern approach to the design of what I shall term *statistical-experimental* investigations. The whole subject has, however, become very complicated as several treatments of one kind have been included, and treatments of several kinds. Thus, experiments may be done with various quantities and combinations of several kinds of fertilizer on several varieties of wheat. Further complication arises when experiments are done on different farms and in different years, and it is necessary to consider to what extent results obtained on one group of farms in one year apply to other farms and other years.

In spite of the fact that sound methods are available, experimenters continue to work with variable material on non-statistical lines, and they get discordant results which they cannot fit into a system. Different workers sometimes get different results in the same subject, and controversies arise. When, in such circumstances, the experimenters turn to sound methods of statistical analysis, involving proper experimental arrangements, difficulties of these kinds tend to disappear. Then, experiments which were previously done on an inadequate scale are increased in size, often they are designed more economically than before, and the advancement of knowledge is made more orderly and certain.

Statistical methods are often regarded as applying only to very large numbers of observations, but that is no longer true. It would be far too costly to replicate some experiments hundreds and thousands of times, and statisticians have had to make do with small numbers. They have, however, developed the theory of errors to apply to small samples as well as to large ones.

There are many chance events that occur in life, to which the general theory of random errors may in some degree be applied.

For example, many telephone subscribers have access to one trunk line, and a multitude of causes determine how many will want to use it at any given instant, i.e. it is to some extent a question of chance whether more than one subscriber will want to use it at once and thus cause delay. In so far as this is true, the extent of delays of this kind can be calculated from the theory of probability which is the basis of the theory of errors. This is typical of a number of congestion problems that arise in telephony, in road and rail traffic, and so on; and although many of them are difficult mathematically, the theory is being applied.

Accidents have a large element of chance in their causation—the circumstances preceding a 'near shave' often differ by only a hairbreadth from those preceding a catastrophically fatal accident, and the theory of probability has been useful for studying accident problems in calculating the effects of chance and showing the importance of other factors. The following is an example.

Records were kept of the numbers of accidents that happened during the course of one year to 247 men workers engaged in moulding chocolate in a factory. Some of the men had no accident, some had one, some two, and so on, a few having as many as twenty-one accidents. The data are arranged in a frequency distribution in the first two columns of Table 15. Now we ask: Were all the variations between the men in the numbers of accidents they suffered due to chance, or were there differences between the men in their tendency to have accidents? Were the 42 men who had no accidents exceptionally skilful or just lucky; and were the 22 men who had ten accidents or more clumsy or unlucky? The average number of accidents per man is 3·94, and even if all the men were equally skilful in avoiding accidents, chance would give rise to some variation. It has been calculated from the extended theory of random sampling that this variation would result in the frequency distribution of the last column of figures of Table 15. This is very different from the actual distribution. We may say, roughly, that 5 of the 42

TABLE 15

Frequency Distribution of Men who had Various Numbers of Accidents. Comparison between Actual and Chance Distributions

(Data by E. M. Newbold. Report No. 34, *Industrial Fatigue (Now Health) Research Board*)

Number of Accidents	*Frequency of Men*	
	Actual	*Chance*
0	42	5
1	44	21
2	30	40
3	30	50
4	25	48
5	11	37
6	12	23
7	15	13
8	8	6
9	8	3
10–15	19	1
15–21	3	...
Total .	247	247

men with no accidents were lucky and the remaining 37 skilful; that one of the 22 men with ten or more accidents was unlucky and the remainder clumsy. Comparisons of this kind between actual and calculated chance distributions have led to investigations that have shown how people differ in 'accident proneness', i.e. in their tendency under given circumstances to suffer accidents. The chance distribution given in Table 15 is calculated by assuming a very simple system of chance variations; more complicated systems taking into account variations in accident proneness have been used in the more advanced investigations on the subject.

CHAPTER VIII

STATISTICAL LAWS

THE central problem of statistics is dealing with groups variously described as collections, crowds, aggregates, masses, or populations, rather than with individual or discrete entities; with events that happen on the average or in the long run rather than with those that happen on particular occasions; with the general rather than with the particular. A fuller consideration of this aspect of statistics is the subject of the present chapter.

Again I shall use the language common in statistical writings and refer to *populations* of *individuals*. The population is regarded in Chapter VI as something from which samples are taken, but here as an aggregate of individuals, which will in most instances be represented by a sample, i.e. I shall not distinguish between the population and the sample.

The population has characteristics and properties of its own, which are essentially derived from and are an aggregate of those of the individuals, although the two sets of properties may be different in kind. In the population, the individuals merge and their individuality is dissolved, but from the dissolution rises a new entity like a phoenix from the flames. The population is at the same time less and more than the totality of the individuals.

This conception is not peculiar to statistics. Rousseau, for example, distinguishes in *The Social Contract* between the General Will and the wills of all the people:

'In fact, each individual, as a man, may have a particular will contrary or dissimilar to the general will which he has as a citizen. His particular interest may speak to him quite differently from the common interest.'

'There is often a great deal of difference between the will of all and the general will; the latter considers only the common interest, while the former takes private interest into account, and is no more than a sum of particular wills: but take away from these same wills the pluses and minuses that cancel one another, and the general will remains as the sum of the differences.'

The general idea is expressed in another way in the following passage from *Old Junk* by Mr. H. M. Tomlinson:

'His shop had its native smell. It was of coffee, spices, rock-oil, cheese, bundles of wood, biscuits and jute bags, and yet was none of these things, for their separate essences were so blended by old association that they made one indivisible smell, peculiar, but not unpleasant, when you were used to it.'

The loss of individuality results from the method of the statistician in confining his attention to only a few characteristics of the individuals and grouping them into classes. Consider a married couple, say Mr. and Mrs. Tom Jones. As a couple their individuality consists in a unique combination of a multitude of characteristics. Mr. Jones is tall and thin, is aged 52 years, has brown hair turning grey, and is a farmer. Mrs. Jones is called Mary and at 38 years is still handsome; she is blonde and is really a little too 'flighty' for a farmer's wife. The couple have been married for 16 years and have three children: two boys aged 14½ and 11 years, and a girl aged 2. In addition to these and similar attributes the couple have a number of moral and spiritual qualities that we may or may not be able to put down on paper. It is by all these, and a host of other qualities that their relatives and neighbours know Mr. and Mrs. Jones; the uniqueness of the combination of qualities is the individuality of the couple.

The statistician who is investigating, say, the ages of husbands and wives in England and Wales is interested only in the ages, and does not wish to describe even these accurately. So he puts our couple in that class (Table 8, p. 51) for which the age of the husband is 45–55 years and that of the wife is 35–45 years. Mr. and Mrs. Jones are now merely one of a group of some 320,000 other couples, and are indistinguishable from the others in their group.

Statistical investigations are not always confined to one or two characters of the individuals, and elaborate methods have been developed for dealing with many attributes, e.g. the ages of married couples at marriage, income, number of children, fertility of the grandparents of the children, and so on, but however many attributes are included, they are very few compared with the number that make up the individuality of each couple.

A population of individuals is the most characteristic and simplest chance system the statistician has to deal with. We do not know, or do not take any account of, the causes of the differences between the individuals, and so we dismiss them as being due to chance, and fasten our attention on the population.

Statistics is essentially totalitarian because it is not concerned with individual values of even the few characters measured, but only with classes. However much we analyse the data to show the variation between the parts, we still deal with sub-groups and sub-averages; we never get back to the individuals. In studying the death rate of a country, for example, we may decompose the general average into sub-averages for the two sexes, for the separate age-groups, for different localities, industries, and social classes; but the death rate of an individual has no meaning. When we think of variation, we think of a mass of variable individuals rather than of one or two being very different from the remainder.

We have already noted in Chapter III that this part of statistical technique in selecting only a few characteristics for investigation, and in classifying the data, is not only necessary because of the limited power of the human brain to apprehend detail, but is a part of the general scientific method. It is an essential step in the development of general scientific laws. However much we know of Mr. and Mrs. Jones in particular, if we know nothing more we have no basis for drawing conclusions about married couples in general. It is only by paying attention to such features as individuals have in common with others that we can generalize. Individuals are important, as such, to themselves, to their neighbours and relations, and to professional consultants—the parson, the doctor, and the lawyer; they have no importance for the statistician, nor indeed for any scientist, except that they, with a host of other individuals, provide data.

Our first and, for most of us, our only reactions to our environment are individualistic. We *are* individuals, our experience is mostly with individuals, and even when considering a group we are conscious mostly of our personal relationship to it. The concept of the population as an entity does not come easily, and our ordinary education does little to correct this defect. The mental effort required to realize this concept is perhaps something like that necessary to appreciate a fugue with its contrapuntal pattern, as compared with the ease of following a tune with simple harmonies.

The characteristics of the population are described by frequencies and by the statistical constants and averages already described, but it is apparently so difficult to think of the reality behind these constants—the mass of individuals—that we personify the population and speak in such terms as 'the average man'. This is only possible because of a similarity between some of the measures of a population and those of an individual; the average height of a group of men is expressed in feet and inches, just as the height of one man is; but the similarity is only superficial.

We have already seen the inadequacy of the average as a description of variable material (Chapter V), but the average individual sometimes is also a rather absurd figure. In 1938, for example, he was among those comparatively rare individuals who died at the age of 58 years. His age in England and Wales in 1921 was 29·9, and in 1938 it was 33·6 years; i.e. in 17 years the average man aged by only 3·7 years! The average family can have fractions of a person. Books on the upbringing of babies usually contain a curve showing the growth in weight of an average baby; but few actual curves are like that. The curve for a real baby may be above or below that for the average and it may have a different slope in various parts. It will also usually have 'kinks' due to teething troubles and minor illnesses, whereas the curve for the average baby is fairly smooth; this paragon among children has no troubles!

Variation is, of course, an important characteristic of populations that individuals cannot have. I have already been at pains to describe this in Chapter V, and to point out how, for example, the deviations from any relationship shown by a contingency or correlation table are as characteristic of the data as the relationship itself. Indeed, without variation, a collection of individuals is scarcely a population in the statistical sense. A thousand exactly similar steel bearing balls (if such were possible) would be no more than one ball multiplied one thousand times. It is the quality of variation that makes it difficult at first to carry in mind a population in its complexity.

All the special properties of populations I have considered arise in aggregates of independent individuals, but there are additional characteristics due to interactions between individuals. The behaviour of men in the mass is often different from their behaviour as individuals. Some men affect (or 'infect') others and such phenomena as mass enthusiasms and panics arise. We speak of mass-psychology. Similarly the effect of an infectious disease on a community of people in close contact is different from its effect on a number of more or less isolated individuals. Statistical description can take account of interactions between individuals, but it is seldom necessary to do so.

Although the individuals in a population vary, the characteristics of the population itself are very stable. Sir Arthur Eddington has well said: 'Human life is proverbially uncertain; few things are more certain than the solvency of a life-insurance company.' This means that we do not know when any individual will die, but an insurance company can estimate the incidence of death in its population of policy-holders with great accuracy.

This contrast between individualistic variability and statistical stability, and the fact that the latter emerges from the former, this apparent paradox of order coming out of chaos, has from time to time given rise to metaphysical speculations. People in the eighteenth century, accustomed to considering the variations between individuals, seem to have been struck by the statistical regularities and saw evidences of a Divine order. Sir Arthur Eddington, on the other hand, presumably taking for granted the regularity of the laws of physics, is more struck by the compatibility with these laws of the unpredictable variation in the behaviour of individual electrons, and offers comfort to those who want to believe in free will and scientific law at the same time. The practical statistician may accept it as a fact requiring no special metaphysical explanation, that mass regularities can often be discerned where the individuals apparently follow no regular laws.

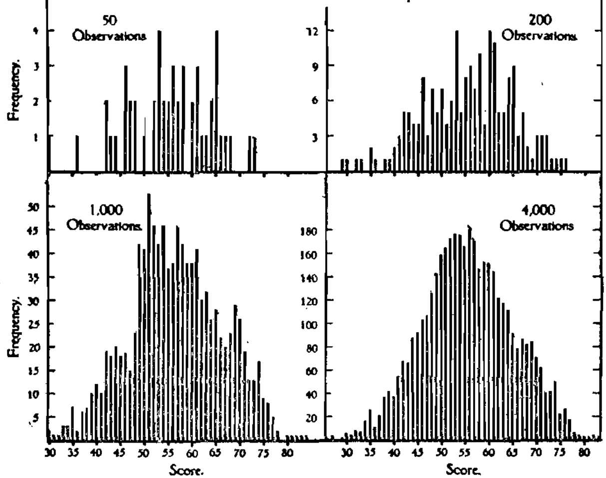
Galton writes of the regularity of form of the frequency distribution in the following terms:

'I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the "Law of Frequency of Error". The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshalled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along.'

Let us re-examine the data from the sampling experiment described in Chapter VI and see if we can repeat Galton's experience and recapture something of his mood.

I have extended the experiment to obtain 4,000 scores altogether. The first thirty are given in the top part of Table 14 (p. 82) in the order in which they occurred, and these together with the 3,970 other scores are the 'large sample of chaotic elements'—and chaotic they undoubtedly appear. I then proceeded to marshal the scores in the order of their magnitude by forming a frequency distribution, and stage by stage stopped to look at the result as the distribution began to grow. The results for 50, 200, 1,000, and 4,000 scores are in Figure 17. Since the scores are whole numbers, I have not grouped them into sub-ranges; the scales of the distributions in the vertical direction have been reduced as the numbers of scores have increased. At 50 scores, there is no sign of any regularity or form in the distribution, but at 200 scores, a vague suggestion of a form seems to be

Figure 17.
Frequency Distributions of Various Numbers of Observations from the Same Population.



emerging; the scores show a slight tendency to pile up in the middle of the range. At 1,000 scores, the form is clearly apparent, although irregularities are still pronounced; but at 4,000 scores, the 'most beautiful form of regularity' is there, almost in perfection. It is not difficult to imagine the regularity that would be apparent were the sample so large as to be indistinguishable from the population.

The formulae and laws that describe populations and their behaviour as opposed to individuals are termed *statistical laws*. The various statistical constants mentioned in Chapter V are elementary statistical laws. Other laws of a higher order of complexity describe how populations change with time or place, or other circumstances. Laws of heredity, for example, are a way of describing how some characters in populations of plants or animals change from generation to generation.

Some statistical laws are discovered by simple observation of the population as a whole. For example, the change in the death rate for the country may be recorded from year to year, without any consideration being given to the changes in the chances of death from various causes, to which the individual is exposed. A public lighting authority could compare two batches of electric lamps by counting how many of each are burnt out after having been in use for, say, 500 hours. Or a colony of the banana fly may be kept in a bottle under standard conditions, and the growth in numbers observed. However, there is nothing necessarily statistical in the technique applied in such experiments, although investigations of this character are often classed as statistical in the widest sense of the word. The introduction of the concept of pieces of matter as populations of electrons or atoms does not necessarily turn an ordinary physical investigation of the macroscopic properties of matter into a statistical one.

Statistical methods and calculations are involved, however, when the laws for the population are deduced from those for individuals. The calculation of statistical constants is a case in point, and the estimation of some quality of a batch of electric lamps from calculations made on the full frequency distribution of lives is another. Estimates, made by demographers, of the size and age composition of the future population from a consideration of the characteristics of the present population and the various birth and death rates, are an important example of the statistical deduction of statistical laws. Such calculations may involve complicated mathematics.

It is implicit in all I have written that statistical laws have nothing to do with individuals. It is no exception to the statistical law that old men have old wives, on the average, if one old man of one's acquaintance *has a young wife. A failure to recognize the distinction* between the two types of laws sometimes leads to attempts to apply statistical laws to individuals, with paradoxical results.

We now return to the starting-point of this chapter—a consideration of individuals. They in the aggregate are the population, and from their characteristics we can calculate those of the population. We cannot perform the reverse process. Individuality is lost, as far as the statistician is concerned, for good and all. Does this mean we know absolutely nothing of the individual when we know the population? Not quite.

Consider a single electric lamp taken at random from the batch represented by the distribution of Table 5 (p. 41). Even if we do not know its life, we know that it will be an exceptional lamp if its life is greater than, say, 2,800 hours—it will be one of 4/150ths of the batch. Indeed, it is more likely to be one of the 89/150ths of the lamps with lives between, say, 1,000 and 2,000 hours.

We are used, in ordinary life, to dealing with data of this kind by introducing the concept of probability. In the example quoted we would say that the probability of any one lamp having a life greater than 2,800 hours is $4/150 = 0{\cdot}027$, and that the probability of the life being between 1,000 and 2,000 hours is $89/150 = 0{\cdot}593$.

This is an application of what is commonly regarded as the statistician's definition of probability as a ratio of frequencies. Corresponding to any frequency distribution there can be calculated a whole series of probabilities of a random individual lying within various stated limits, and statistical probability is a device (a verbal trick!) for attaching to the random individual the characteristics of the whole distribution. In this way, a population is epitomized in an individual much more satisfactorily than in the concept of 'the average man'. But statistical probability does more than this. It corresponds closely to the more popular idea of probability as a measure of the strength of belief in a thing. Most people if asked what is the probability of a tossed penny falling heads uppermost would reflect that heads was as likely as tails and would reply: one-half. The statistician, if in a pedantic mood, would reply: in the hypothetical population of tosses, one-half of the total give heads, therefore the probability of a head is one-half. An alternative method of expression is to state that the chances of a head are even, or for the lamps, that they are 593 to 407 in favour of a life of between 1,000 and 2,000 hours.

Probability is, in ordinary life, also applied to events that do not occur as frequencies. We speak of the probability of or the chances in favour of a particular horse winning a race. Even in such instances, however, I think that people carry at the backs of their minds the idea of frequencies; they in effect imagine a lot of races, in a given proportion of which the particular horse wins. The idea is described in the following quotation from a lecture given by Karl Pearson in 1892:

'A friend is leaving us, say in Chancery Lane at 4 o'clock in the afternoon, and we tell him that he will find a Hansom cab at the Fleet Street corner. There is no hesitation in our assertion. We speak with knowledge, because an invariable experience has shown us Hansom cabs at 4 o'clock in Fleet Street. But given the like conditions within reach of a suburban cab-stand, and our statement becomes less definite. We hesitate to say absolutely that there will be a cab: "You are sure to find a cab", "I believe there will be a cab on the stand", "There is likely to be a cab on the stand", "There will possibly be a cab on the stand", "There might *perhaps* be a cab", "I don't expect there will be a cab", "It's very improbable", "You are sure not to find a cab", etc., etc. In each and every case we go through some rough kind of statistics, *once* we remember to have seen the stand without a cab; on occasions few and far between, "perhaps on an average once a month", "perhaps once a week", "every other day", "more often than not there has been no cab there". Certainty in the case of Fleet Street passes through every phase of belief to disbelief in the case of the suburban cab-stand. If once a month is the very maximum of times I have seen an empty cab-stand, my belief that my friend will find a cab there to-day is far stronger than if I have seen it vacant once a week. A measure of my belief in the occurrence of some event in the future is thus based upon my statistical experience of its occurrence or failure in the past.'

Thus probability in its most general use is a measure of our degree of confidence that a thing will happen. If the probability is 1·0, we know the thing will certainly happen, and if the probability is high, say 0·9, we feel that the event is likely to happen. A probability of 0·5 denotes that the event is as likely to happen as not, and one of zero means that it certainly will not. This interpretation, applied to statistical probabilities calculated from frequencies, is the only way of expressing what we know of the individual from our knowledge of the population.

Statistical laws, which describe the characters and behaviour of populations in one way or another, may be transformed into probabilities—i.e. from them the probabilities and frequencies in the population may be calculated. Thus, statistical laws are the chance laws referred to in the early part of Chapter VII.

It may have been noticed that probabilities have been calculated from the frequencies of a distribution, either known as for the lamps, or assumed as for the penny. In general, it is necessary to have some data on which to calculate probabilities. I am often asked what is the probability of some queer or interesting event, without being given any data. Statisticians do not evolve probabilities out of their inner consciousness, they merely calculate them.

CHAPTER IX

STATISTICAL REASONING

STATISTICAL facts may have interest purely as a description of something that has happened or of an existing state of things, and certainly have great value in practical affairs. But we are seldom content with this use; we try to interpret the facts and expect them to tell us something of the underlying processes at work in the world we are studying. It is in this connexion that statistics is mostly misused. In this chapter I propose first to illustrate some of the misuses, and then to describe the ideas behind the ways in which statisticians try to learn from statistics. Statistical reasoning is not really different from any other kind of reasoning, and since the statistical method is a special case of the general scientific method I shall devote some attention to the latter.

I think there are two reasons why statistics are so much misused. First, the desire to interpret them is so strong within us that it is almost an instinct, and we are apt to embark on interpretations too easily, without adequate mental preparation, and even without training in scientific habits of thought. Consequently we tend to jump to the superficially obvious conclusions, which all too often are not the correct ones. The interpretation of statistics is a matter for the expert—although one may be an expert without necessarily having a university degree in the subject. Second, statistics are so often made to serve the purposes of a propagandist—the man who does not use figures to arrive at the truth of a matter, but thinks he knows the truth and only wishes to convince other people. The propagandist may misuse statistics honestly, from ignorance, or dishonestly and deliberately; or he may countenance that more subtle form of dishonesty of presenting data in a way that is formally correct, but misleading, leaving it to the public to draw the obvious but false conclusion. This last trick is most insidious when it is accompanied by some remark intended to disarm criticism, such as: 'I know these figures may be interpreted in several ways, but . . .'

One favourite trick of the propagandist is to use some impressive but irrelevant figures to give a spurious appearance of precision under the cover of which a dubious argument is 'slipped across' to the public. Commercial advertisements often do this. 'Five thousand and sixty-seven typists were asked what they prefer in shoes and four thousand nine hundred and ninety-five, or 98·6 per cent., prefer comfort to smartness; therefore buy XYZ shoes.' That is the kind of argument one meets.

A common source of error is the use of inaccurate data, of misleading methods of presentation, or of data that are so incomplete as to be misleading. I have shown in Chapters II to VI many ways in which this may arise. Here is an example about a subject that has been of a very lively concern to the British public. On 13th May 1941 most newspapers reported the following facts of the numbers of German night bombers brought down in raids over Britain:

'The total for the eleven nights of May now stands at 133; the previous highest figure was 90 for the whole of April.'

That suggests an enormous improvement in the effectiveness of British night defences. Indeed, the quantitative impression is of an increase in the number of bombers brought down per night from $90/30 = 3 \cdot 0$ to $133/11 = 12 \cdot 1$. This comparison should not be so interpreted without a knowledge of the numbers of bombers used. Those figures were not given, but we noticed at the time that most of the raiding in April was confined to the ten or eleven nights before that of the full moon. Full moon was on April 11/12th and May 10/11th. Up to and including the night of April 11th, 51 night bombers were brought down, giving a rate of 4·6 per night. This is less than the rate for the first eleven nights of May, but not so much less as the figures originally quoted suggest. We cannot be sure that the number of nights before the full moon is the significant basis of comparison, but it is almost certain that the crude comparison gave an impression which, to the Britisher, was unduly optimistic.

The pattern of cause and effect in the world which produces statistical data is very complicated, and to any set of figures several plausible interpretations are usually possible. It is a common error to consider or give only one interpretation, to the exclusion of others that may be equally reasonable but perhaps less agreeable to the propagandist. The following advertisement appeared in 1931:

'It is men of exceptional experience who are buying X . . . cars to-day.

'87 per cent of X . . . cars to-day are bought by men who have owned six other makes of cars before.'

I suppose it is unlikely that as many as 87 per cent of all makes of cars are bought by such veterans as those mentioned in the advertisement, and the purchasers

X . . . cars are probably exceptional, but they may be exceptional in their fickleness—and do the makers of the X . . . cars wish us to believe that they do not get many repeat orders? Those are possible interpretations of the data. X . . . cars may, for all I know, be as good as most other makes, and better than many; but as a recommendation, the facts of the advertisement are practically valueless.

The following extract from a letter to a newspaper breaks several rules of statistical good conduct:

'An inquiry some years before the last war was made in certain towns all over England, Manchester included, as to the effects of the use of oatmeal among children and in public institutions, and the following is taken from it: in Manchester, 2,333 children in all were questioned. In one school of 200, 84 per cent were regular users, and the teacher stated, judging from regularity of attendance, that those getting oatmeal were the most satisfactory.

'In a girls' school of 182 pupils 33 had porridge, and the headmistress reported:

'The majority of oat-users are strong children, well nourished, and class work good. The non-users are not so strong and more liable to take colds and infectious diseases, and class work only moderately good.'

Other data of the same kind were given, all purporting to show the virtue of oatmeal. Now the essential information in the above extract is the better health of children who use oatmeal, but this is given only in vague qualitative terms. No statistician would rely on such general impressions as are quoted. What were the attendance results of the oatmeal-users as compared with the others? What were their sickness records? Moreover, the poorness of the data is covered over, doubtless unintentionally, by some very exact but irrelevant figures. It does not matter 'two hoots' how many children were questioned, or how many took porridge. Without these figures, the data would be seen to be what they are—weak. However, even if the facts are taken at their face value, the letter-writer errs in considering only one of the factors that could have contributed to the alleged results. Most people take milk with porridge, which might be extra to milk taken otherwise, and that might be the cause of the improved health. All we know, if we know anything from the data, is that oatmeal plus milk plus the condiments are good for health, as compared with the food that is eaten as an alternative.

Wrong conclusions are sometimes drawn from data of quantities that change in time, and it is a standard part of the statistician's functions to recognize 'nonsense correlations'. For instance, Mr. Udny Yule refers to the fact that the proportion of marriages solemnized in the Church of England and the death-rate for the country have for many years been decreasing—there is a correlation between the two quantities. I doubt, however, if anyone supposes that this fact implies a causal relationship, and that a law prohibiting the solemnization of marriages in Anglican churches would reduce further the mortality rate of the nation.

A neglect to consider the errors of sampling or the effects of random fluctuations sometimes leads to false interpretations of statistical data. These effects have been dealt with in Chapters VI and VII.

So much for the things that should not be done in handling statistical data. Let us now consider the e positive aspect of what should be done. To vide a special case, on which the discussion will be tred, I will remind the reader of Lamb's *Dissertation n Roast Pig*. According to Lamb's imaginary inese manuscript the art of roasting, or rather oiling, was accidentally discovered in the following anner. A swineherd, Ho-ti, left his cottage in e care of his eldest son Bo-bo, who, being fond f playing with fire, let some sparks escape into a undle of straw, which reduced the cottage to ashes. Together with the cottage . . . what was of much more importance, a fine litter of new-farrowed pigs, no less than nine in number, perished. Bo-bo was in utmost consternation. . . . While he was thinking what he should say to his father an odour assailed his nostrils. A premonitory moistening at the same time overflowed his nether lip. He stooped down to feel if there were any signs of life in the pig. He burnt his fingers, and to cool them he applied them in his booby fashion to his mouth. Some of the crumbs of the scorched skin had come away with his fingers and for the first time in his life (in the world's life indeed, for before him no man had known it) he tasted—*crackling*! . . . The truth at length broke into his slow understanding, that it was the pig that smelt so, and the pig that tasted so delicious.'

When Ho-ti returned there were at first the mis understandings Bo-bo expected and feared, bu gradually the great truth was borne in upon Ho-ti mind and ultimately 'both father and son fairly s down to the mess, and never left off till they h dispatched all that remained of the litter'.

Ho-ti of course wanted more roast pig, and so, often as the sow farrowed, so sure was the house

Ho-ti to be in a blaze'; and later, after the secret had been dragged into the light in a law court, 'there was nothing to be seen but fires in every direction'.

This is an example of the empirical method. Certain results are observed to follow from a certain set of circumstances, so in order to repeat the result the circumstances are repeated. This method is often regarded disparagingly, but it is widely and successfully used. If we know that certain desired ends can be achieved by certain means, we rightly use those means, without waiting until we can find out how they work and if all the means are necessary to achieve the ends. Our forefathers would have been foolish had they waited for the discovery of vitamin C before making use of the knowledge that fresh vegetables in the diet prevent scurvy. Indeed, medicine is a fine example of the successful use of empirical knowledge (I do not imply that medicine does not also use scientific knowledge).

The empirical method, however, does not take us very far, and often leads us astray. Experience leads us to believe that always, if we can re-establish exactly all the circumstances that gave rise to a result, that result will be repeated exactly; but we can never be sure of re-establishing all the circumstances. Moreover, not all of them are essential, and without some analysis of the causes that operate we may repeat certain non-essential circumstances and omit essential ones. For example, we are not told which way the wind was blowing when Ho-ti's house first burnt; as it happened that did not matter, but if it had, and Ho-ti had not taken it into account, the empirical method might have failed him. He was lucky to have included the important factor in his subsequent trials.

The empirical method is not only uncertain; it is often wasteful. Not only was it wasteful for the compatriots of Ho-ti to burn their houses in order to roast pigs, but the failure to discover the general science of the use of heat for cooking may have involved them in burning their boats to cook fish and might even have led to disastrous experiments into the burning of crops to improve their taste. It is desirable to discover the causes of an observed phenomenon so that the essential factors can be reproduced in the most advantageous way and applied generally.

Finally, it is only when our knowledge of the causes is fairly detailed that we can continue with investigations to improve the result. Until they had discovered that it was the application of heat that cooked the pig, it would scarcely be feasible for the people of Ho-ti's time to experiment with the effects of different degrees of heat and discover the uses of boiling, frying, and so on.

All the foregoing are utilitarian reasons why we are not content with the empirical method, but transcending them all is the intellectual desire we have to 'explain' things: to describe the relations between different happenings: to reduce our knowledge of the universe to as few general principles as possible.

It is the main function of science to analyse the causes of events and build up a system of general laws, and so we regard scientific knowledge as the opposite of empirical knowledge. This is the sense in which I use the word scientific in this book. Science and empiricism, however, differ only in degree, and scientists of the present generation, at least, are very modest in not claiming any sort of finality for the laws they formulate. Ho-ti was working on the very lowest level of empiricism when he burnt his house to roast pig and he would have been more scientific had he recognized that it was only necessary to have some sort of fire. We know now that even fire is not necessary, and that heat of a sufficient degree produced in any way, e.g. electrically, will roast the pig; but it is conceivable that some scientist of the future will discover the essential chemical and physical changes that occur when pork is roasted and will give us other ways of achieving the result. What the next stage after that can be I cannot imagine, but it would be very rash to say there will be no next stage. The point is that none of these stages is purely empirical and none is purely scientific; they differ only in degree.

The scientific method is so efficient, and has been successful in giving man so much power over his environment—power to construct as well as to destroy—that we are apt to overlook its limitations. A scientific description is by its nature a simplified description of a phenomenon and its relation to other parts of the universe, and is far from complete. This has been shown strikingly in connexion with the science of dietetics. Earlier in this century, food values were largely measured in terms of calories and the chemical constituents—fats, carbohydrates, proteins, and so on. Then it was found that these things were not enough, and that a diet was deficient unless it contained a due quantity of the various vitamins. It is because they have no faith in the completeness of the existing scientific description of dietetics that most people prefer to rely mostly on the empirical method in arranging their diet, choosing to eat such natural foods as general experience has shown to be good, and relying on special sources of vitamins only in special circumstances.

I think that considerations of this sort may lie at the root of the so-called mistrust of logic and reason that is sometimes attributed to the English. To mistrust logic and reason is—well, *irrational*; but it is not irrational to lack faith in the completeness of the data on which some reasonable conclusions are based. Such an argument does in some measure justify the Englishman's alleged preference for empiricism or 'muddling through' and for his reluctance to follow general principles to their 'logical conclusion'. It is certainly important in statistical investigations not to forget that the data do not tell everything, and that in their summarized, reduced form they leave out a lot of information that may be important.

Let us imagine how a physicist, working in our modern spirit, but with the knowledge available at the time of Ho-ti, would investigate the phenomenon of the production of roast pig, to discover which of the circumstances surrounding Bo-bo's and Ho-ti's experience were essential. He would first consider, as far as he was able, all the circumstances: the fact that there was fire, that the fire started in straw, that a house was destroyed, that there was a wind blowing in some direction (unrecorded), that the pigs were new-farrowed, and so on. Then, if he was painstaking but unimaginative, the physicist might conduct a series of experiments, reproducing in each one of the circumstances in some way *different from Bo-bo's*. He might put the pig in or over a log fire; kindle the fire in the house with paper instead of straw; destroy the house over a litter of pigs by gunpowder instead of fire; and so on. In this way, he would soon discover the irrelevance of everything but fire and pig, although he might be a bit puzzled because all fires did not produce the desired effect (if they were not hot enough or did not last long enough, or were too hot or lasted too long). This last observation would probably lead him to try cooking the pigs on fires of various sizes and durations—and so the process would go on.

This is the experimental method of investigation: the method of producing the circumstances surrounding a phenomenon in various ways and combinations, but always under experimental control. The art of this method lies in the proper choice of combinations of circumstances and in the craftsmanship of exercising the required control; and if these have been well done there is little difficulty in correctly appraising the results, although great acumen may be necessary to weld them into a coherent scientific theory.

In most fields that are the subject of statistical inquiry, the opportunities for controlled experiments are very few. Society will not permit many experiments on man. So the statistician, debarred from varying the circumstances surrounding the subject of his investigation, has to observe the results of such variations as occur without his intervention and learn from them, disentangling as much as he can from the 'tangled skein' of causes and effects. Thus, if the Chinese authorities had prohibited experiments by the physicist into the production of roast pig, the statistician would be called upon to observe closely the results of the fires that appeared 'in every direction', in the hope that the circumstances surrounding them would be varied enough to enable him to decide which were essential and which were not.

This limitation does not preclude the application of scientific methods, as some scientists suppose. How much experimenting can the astronomer and meteorologist do? Are they unscientific? Surely it does not matter, in principle, whether the variations in circumstances that are studied are produced by the investigator, or not. The important thing is that sound reasoning should be applied to the interpretation of the results.

The method of inferring laws from observations on systems that result from uncontrolled variations is to many people *the* statistical method. Indeed, Mr. Yule has defined statistical methods as 'methods specially adapted to the elucidation of quantitative data affected by a multiplicity of causes'.

There are three main ways in which the statistician treats observations so as to separate the effects of the various causes. Sometimes (unfortunately not very often) he can observe two things that differ to an important degree only in respect of the one factor under investigation; and then a straightforward presentation of the results tells us everything about its effect. The second way is to correct for the effects of disturbing factors, just as we corrected for the difference in age composition when comparing the death rates of the two districts of Wales (p. 73). The third way is to eliminate by averaging the effects of disturbing factors that operate in various directions. For example, if we wish to measure the effect of town as compared with country conditions on the death rate, we might first measure the rates for a number of towns, correct for age and sex distribution, and then average the results, expecting that the variations between towns due to other factors such as climate and locality will average out. Then we might do the same for a number of country areas. The difference between the two averages would then be regarded as measuring the required effects.

The use of correlation methods is an extension of this idea. As I have already stated, the trend shown in Figure 15A (p. 53) discovers the average effect on the wheat crop of changes in the area cultivated, and the diagram is arranged to emphasize this effect. A correlation may result directly from the operation of a cause, but I have already emphasized that its existence does not prove a causal relationship. A careful analysis is necessary before such an interpretation of a correlation is legitimate, and usually there must be additional grounds to support it. In this analysis, statistical methods play a part, particularly the method known as *partial correlation.* The following is an example.

In an investigation by Mr. D. V. Glass into factors associated with changes in birth rates, the following data for 1930–32 for the separate counties of England and Wales were used: (1) the gross reproduction rate, which is similar to the net reproduction rate (p. 61) except that no account is taken of the incidence of death; (2) the percentage of females over 15 years of age who were unmarried, which I shall call the percentage spinsterhood; and (3) the percentage of females over 15 years of age who were in employment, which I shall shortly describe as the employment rate. These three quantities were taken in pairs and correlated (see Chapter IV and p. 76), with the following results:—(*a*) there was a correlation coefficient of −0·433 between factors (1) and (2), expressing a fairly weak but definite tendency for the reproduction rate to decrease as the percentage spinsterhood increases—

a result not unexpected; (*b*) the correlation coefficient between factors (1) and (3) was $-0{\cdot}625$, expressing a stronger tendency for the reproduction rate to decrease as the employment rate among women increases; (*c*) there was a correlation coefficient of $+0{\cdot}530$ between factors (2) and (3); i.e the greater the percentage spinsterhood the greater is the employment rate among women. On the face of things, it looks as though a high percentage spinsterhood and a high employment rate both reduce the reproduction rate; but these two quantities are not independent, and their effects are intermingled. Suppose, for example, the employment rate of itself had no real effect on the reproduction rate; it would reflect in some degree the influence of the percentage spinsterhood and show an *apparent* effect, on the argument:

high percentage spinsterhood leads to low reproduction rate,
high employment rate is associated with high percentage spinsterhood,
therefore, high employment rate is associated with low reproduction rate.

This effect could explain some, at least, of the apparent correlation between factors (1) and (3); and, similarly, any causal effect the employment rate had could explain some of the apparent correlation between factors (1) and (2). The method of partial correlation enables us to separate out these effects. The *partial correlation coefficient* between factors (1) and (2), which measures the effect of percentage spinsterhood alone, is $-0{\cdot}15$, expressing a very weak association, which is of negligible importance. That is to say, if the employment rate among women is kept constant, the percentage spinsterhood is practically unrelated to the reproduction rate. The partial correlation coefficient between factors (1) and (3) is $-0{\cdot}52$, so that if the percentage spinsterhood is kept constant there is an appreciable tendency for a high employment rate among women to produce a low reproduction rate. These results, as far as they go, suggest that, in order to increase births, it is not much good increasing marriages, but the discouragement of employment among women might have some effect. It is necessary in giving this summing up of the results to emphasize the words *as far as they go*. The results only apply to such variations as occurred between counties in 1930–32, and there may be other important causal factors, of which no account has been taken.

I have described the exhaustive investigation of all the circumstances surrounding the first production of roast pig as sound but unimaginative; progress in science has been much facilitated by the imaginative procedure of using working hypotheses in planning and making investigations. I cannot discuss this method of approach to a problem in full, but roughly it consists in making a tentative hypothesis based on existing knowledge and ideas, and testing it by arranging experiments that give one result if the hypothesis is correct, and another if one or more of a number of alternative hypotheses are correct. For example, the Chinese physicist, on being told that Bo-bo and Ho-ti both burnt their fingers when they first touched the pig, might, in a flash of genius, see that heat had something to do with the transformation of the pig, and would direct his experiments to testing the hypothesis that fire was the only factor of importance. As admissible alternative hypotheses, he might allow the possibility that the important factors were combinations of the nature of the combustible material, the direction of the wind, the age of the pigs, and so on. Then if he roasted pigs in two fires, in which the only common factors were fire and pig, and in which all the other factors differed, the physicist would at one blow test his hypothesis against the alternatives. If the pig was always roasted, his confidence in the hypothesis would be increased, although he could never be sure of it, because he could never be sure of having tested it against all possible alternatives. If the hypothesis was proved false he would have to think again, and, from his observations on the experiments already made, would form some new hypothesis which would in turn be tested. Used in this way, hypotheses are guides to experimental strategy.

In the early days of a new science there is little or no knowledge to form the basis of working hypotheses, and the patient collection of facts constitutes the main activity. Very soon, however, speculation on the meaning of the results begins, and tentative theories are developed. Then experiments and observations are made to test the theories, which are further developed, and modified or remade as a result of the new experience gained; and again the new theories are put to the test, and so on, experimental test alternating with theoretical speculation.

These two phases do not always proceed in the best order. I think, for example, that the early work of the biometricians (see p. 169) had biological significance chiefly in providing data, and that it failed to develop because it did not go far enough beyond this stage. In economics, on the other hand, there has been an apparent tendency for theoretical speculation to outstrip verification by observation of the real world (p. 167).

The use of working hypotheses, both the main one and its alternatives (the alternatives are all too often neglected by the amateur), is very important in statistical research. They guide the statistician in planning his inquiry, in choosing what data to collect or use, in arranging and presenting them, and in deciding what statistical constants to calculate; and finally the results are examined in their light. Without such aid in selecting from the enormous range of possibilities, progress in knowledge would be slow, and much effort would be wasted in useless work. For example, in investigating the porridge question mentioned in the letter quoted on page 125 we might dismiss the likelihood of the condiments having any effect on health, and adopt the hypothesis that oatmeal and milk both have good effects, with the three alternatives that benefit is derived from (1) oatmeal alone, (2) milk alone, and (3) neither milk nor oatmeal. Then we would measure separately the health of children who took (*a*) oatmeal and milk, (*b*) oatmeal without milk, (*c*) milk alone, and (*d*) neither milk nor oatmeal; if it were impossible to find anyone who took oatmeal alone, an experiment might be necessary. If the question of the effect of the condiments was also included, the inquiry would be more complicated. In making these suggestions, I have neglected the important questions of the quantities of oatmeal and milk given to the children, and of the alternatives to these foods (for presumably children who do not eat porridge have something for breakfast); all these would need to be considered in a comprehensive study.

Thus, it is better to approach an inquiry in the light of existing knowledge, and to arrange it to answer certain specific questions, than to collect the data, subject them to a routine of statistical reduction, and then passively accept whatever results emerge. I do not deny that this passive approach may often yield good results, but it is inferior, except in some new field where there is no previous knowledge on which to base hypotheses. The use of statistical data to prove a case, in the sense of demonstrating it, is unscientific; but their use to prove a case in the old-fashioned sense of testing it is scientific and profitable. A statistical inquiry should be approached with a mind that is open but not empty.

The success of this method of approach depends to some degree on the main working hypothesis being somewhere near the truth. A false hypothesis can do no permanent harm, for ultimately it will be discredited, and investigations inspired by such often lead to valuable discoveries. Nevertheless, a false trail may for a time be set and time may be wasted; and an investigator who was too often on the wrong trail would not get very far.

There are no golden rules for the formulation of hypotheses, and their quality and success depend much on the knowledge and experience of the investigator in the field in which he is working, and on his intuition, acumen, and genius. Hypotheses may grow in the investigator's mind in the course of his work, they may come in a mental flash, or they may be suggested by some external accident. Some workers find it helpful to write individual facts on separate cards and play a kind of game of 'patience' with them, sorting the cards into combinations suggesting a variety of relations between the facts. The ability to formulate fruitful hypotheses and design experiments to test them is the quality of a first-rate scientist. In addition to this personal quality, habits of thought and even prejudice have their influence on the kinds of hypothesis that will be entertained. For this reason, impartiality is essential; and an investigator is most likely to be impartial if he is disinterested in the issue of the inquiry. The investigator should not be narrow-minded, and should be prepared to consider any reasonable alternatives to the main hypotheses he favours, but he cannot afford to waste his time on unreasonable ones. Sidney and Beatrice Webb, who have much of interest and value to say on the subject of social investigations, have written:

'We have found it useful, in the early stages of an investigation, deliberately to "make a collection" of all the hypotheses we could at that stage imagine which seemed to have any relevance whatever to the special kind of social institution that we were dealing with. We noted them all down on our several sheets of paper, and others as we went along: wise suggestions and crazy ones, plausible theories and fantastic ones, the dicta of learned philosophers and those of "cranks" and monomaniacs, excluding those that we thought had no possible relevance to our work, such as the prophecies extracted from the measurements of the Great Pyramid, or those of the astrologers.'

This passage suggests a breadth of outlook which I imagine most scientists would applaud, but not so many show; but even the Webbs stuck at taking astrology seriously. Yet I cannot see why they should exclude this and include the theories of cranks and monomaniacs; those who believe in astrology have a perfect right to say that, in this matter, the Webbs show prejudice. The exact position at which one draws the line between what is reasonable and unreasonable is largely a personal matter.

Hypotheses formulated at any one time also tend to follow trends or fashions. For example there are great similarities between the theory of natural selection and the *laissez-faire* theories of economics, and there is some vogue in these days for applying the dialectical process to the pursuit of knowledge in various fields. The favourite hypothesis with which statisticians usually first examine data is, that the observed variations and effects are due to random errors or to chance rather than to the operation of newly discovered causes. This can be tested by the theory of errors, and the statistician will almost invariably hold it as long as it is compatible with the data. In this way, one of the statistician's chief functions is to act as a devil's advocate against the admission of new knowledge. It has been said that 'Bacon was eminently the philosopher of *error prevented*, rather than of "progress facilitated".' The same might almost be said of the statistician. In the fields to which statistics is mostly applied, the prevention of error is a most necessary function; whereas there are plenty of people ready to facilitate progress.

Frequently, several reasonable hypotheses are compatible with the data, and then fresh data are necessary before any discrimination can be made between them. Even without such data, however, one hypothesis may be preferred to the others. The hypothesis already referred to, of chance being the cause of the effects, is a favourite one, and derives from the scientist's general preference for the simplest explanations involving the introduction of the fewest new quantities or ideas. Which of the alternatives is the simplest, depends on the investigator's idea of the general scheme of things, for the simplest is that which fits most easily into such a scheme. For example, data have been given showing that the severity of attack from smallpox tends to increase as the time elapsing since vaccination increases, and is greatest in patients who have not been vaccinated. To those who are not against vaccination on other grounds, the obvious inference is that this treatment is effective as a protection against bad attacks of smallpox. Anti-vaccinationists, on the other hand, find it easier to explain the above data on theories which, to the outsider, seem very complicated. As a statistician I cannot condemn those theories, although as a man who generally favours orthodox views in science I prefer the more obvious inference.

Thus the statistical method, like scientific method in general, is based on certain fundamental principles, but it is not entirely automatic in its operation, and progress in knowledge depends to a considerable degree on the personal qualities of the investigator. He must be creative of ideas, yet should strike a nice balance between being too far-fetched and fanciful on the one hand, and being so conservative on the other that he impedes progress by his unwillingness to admit new knowledge and ideas.

I have insisted that the statistical method of investigation is scientific. Its critical apparatus is sufficiently well developed and discriminative to prevent an undue proportion of false conclusions being reached as a result of statistical inquiries (a certain amount of risk must be taken if progress is to be maintained). At the worst the verdict may be that a particular inquiry teaches nothing new. But the question remains: Is the method powerful? Do statistical investigations often lead to positive conclusions? In the next two chapters I give an estimate of the usefulness of the statistical method in the various fields of application, but here give only a brief general answer to the questions.

When clear conclusions emerge obviously from a simple arrangement of the data the statistical method is useful in a positive way. When, however, the pattern of causes and effects is complicated, and elaborate statistical analysis is necessary, positive conclusions are not often reached. So often, several hypotheses are compatible with the data, and when an analysis like that described on pages 134 to 136 is performed one cannot be sure that it is complete and that all factors have been accounted for. Consequently, the results of a purely statistical inquiry do not usually rise much above a fairly low level of empiricism, and any more scientific laws that are arrived at are largely justified on theoretical grounds.

I present this view somewhat in a spirit of disillusionment. When first introduced to me, the methods of statistical analysis, particularly that of partial correlation, seemed to have unlimited power to penetrate the secrets of nature. I think, too, that this enthusiasm inspired the statisticians who developed the methods during the early years of this century and has been shared by many others, although I have no documentary evidence of this. Certainly, compared with such high hopes, the achievement has been disappointing.

Are statistical methods therefore useless and must we abandon them? No! They *are* powerful, even if their power is limited, and there are fields in which only they can advance knowledge. We must persevere with them in a spirit that is steadfast, if somewhat chastened, believing that progress will be maintained, if more slowly and with greater difficulty than once seemed likely. Moreover, statistical methods are proving exceedingly powerful and are achieving much in the statistical-experimental kinds of investigations described in Chapter VII (pp. 103 to 105).

The use of statistics for discovering the forces at work in the social, economic, and similar spheres, where experiments are impossible, is a very difficult application of the scientific method. Many causes and effects are entangled so that it is hard to separate and relate them. Yet even the ordinary citizen needs to have some ability at least to distinguish what may be truth from what is probably falsehood, especially in a democracy, where he has to make up his mind on many difficult public questions and contribute to the growth of public opinion. Surely it is an important task of education to give the citizen this ability by teaching the elements of statistical reasoning. If this is done, people will develop not only the ability to look at controversial social and similar problems scientifically and dispassionately, but also the habit of doing so.

CHAPTER X

STATISTICS IN AFFAIRS

THE use of statistics in the business of running the country through its political, commercial, and social institutions—in those activities that determine the health, wealth, and happiness of mankind—is the oldest and the most considerable use. The Ancient Egyptians had a centralized form of government administered with the aid of systematic statistical knowledge of the economic conditions of the country (e.g. regular returns were made of the level of the Nile, on which the prosperity of the country so much depended). The English Domesday Book contains the results of a statistical survey, and there are evidences of statistics having been used in administration now and again during the subsequent centuries. There was in England a quickening of interest in commercial statistics early in the eighteenth century, but that only heralded a dawn which broke early in the nineteenth century. The importance of this aspect of statistical activity to-day is shown by the fact that most people seem to regard it as the whole content of statistics.

The need for statistical knowledge in running a concern increases as the concern becomes larger and more complicated. One man can conduct the affairs of a family or a small business with few figures, but as the scale of the enterprise becomes larger it becomes less possible for one man, or even a few men, to have at the same time the necessary intimate knowledge of all the parts and the broad knowledge of the whole. Hence an organization is set up whereby the men in central positions work largely through statistical knowledge of the parts under their control—knowledge that is statistical in the two senses of being numerical and summarized. As more industries have fallen under the control of large combines, and more of our activities have come under the control of the largest combine of all, the State, statistical knowledge has become increasingly important. Planning is the order of the day, and without statistics planning is inconceivable.

In most of the applications of statistics considered in this chapter the most elementary statistical methods suffice. Consequently, this aspect of the subject is comparatively uninteresting to the mathematical statistician who specializes in the development of theory and technique; but it is well that it does not require great mathematical knowledge, at least to follow these parts of statistics which, after all, are of the greatest social importance. The economical and expeditious handling of masses of figures in large concerns is, however, a skilled job, requiring powers of organization and a knowledge of what can be done with the aid of the very expensive and intricate accounting and sorting machines now available.

First let us see how statistics are used in running things after policy has been decided. Their most elementary use in administration is in the balancing of the activities of one part of a system against those of another, to secure that supplies equal requirements, and that there are no 'bottle-necks' or parts that are not employed to the full.

For the national government, the necessary statistics range from extensive figures of the expenditure of the various departments of state and the yields of various taxes, required by the Chancellor of the Exchequer in framing his budget, to the numbers of boots of various sizes that will be used by the army, required by the stores department of the War Office in making requisitions. War much increases the need for figures of these kinds. The allocation of labour, shipping, food, raw materials and resources of all kinds, which in normal times is left to the free play of economic forces, is now (in 1942) directly controlled by Government officials working in the light of statistics of the resources available and the requirements of the various services, industries, districts, and so on. There is little I can say about this function of statistics, but a few minutes' consideration will convince readers of its vital importance.

One important subject is the measurement of the national income. That, rather than the revenue of the Government, determines the resources available for war or other national purposes. For many years, the estimation of the national income of Great Britain has been left largely to the initiative of private individuals, but since 1941 the Government, with its access to official records and all its resources, has taken a hand, and has started to publish annually estimates of the national income and expenditure.

Local authorities need statistical information to enable them to adjust their supplies of various public services to the needs, both immediate and future, of the districts they serve. When building a new housing estate, for example, water, gas, electricity, sewers, schools, transport, and so on have to be provided, in quantities that are sufficient but not excessive. Such legislative measures as the raising of the school leaving age involve problems such as making decisions as to how many new schools and teachers will be required, and to solve these recourse must be had to census data.

The planning of production in a large factory or combine is now a part of what is known as Scientific Management, and many firms now have a planning department to co-ordinate the activities of the other departments. The requirements of the sales department with their delivery dates are translated into orders to the various production departments, with intermediate delivery dates so arranged that the final products are delivered as required; these orders are translated into orders for raw materials, tools, and labour, and the whole activity is organized and timed so that the work flows without interruption. Moreover, to secure efficiency, it is necessary as far as possible to balance the sizes of the various departments so that they are large enough to meet all demands made upon them and yet are not unnecessarily large. For this work, statistical returns and charts are much used. The whole subject is highly developed and involves special knowledge and experience, although the statistical methods used are not very elaborate.

Statistics is useful in administration in providing measures of performance and efficiency. Balance sheets and statements of accounts (not necessarily the published ones) have this use. Various special indexes are also devised, such as the ton-miles of freight carried per engine-hour, used by the railways, and the ratio of management expenses to premium income, used by insurance companies. In a factory, the average output per man-hour or per machine-hour and the percentage of the total materials wasted or spoilt may be useful quantities. With the aid of figures of these kinds, the operating efficiency of a concern can be compared from time to time or from one section to another; one firm can be compared with another or with the average for all firms in the industry; and salesmen can be compared. The data do not state the causes of inefficiency, if any, and much less do they show what to do to effect improvements. They are merely pointers, showing administrators where things are going well and where improvement may be sought; and their value depends entirely on the use made of them.

A striking example of the use of statistical records for increasing efficiency is provided by a costing system set up by the L.M.S. Railway Company for their locomotives. Expenditure on locomotives represented such a large item in the expenses of the company that it was considered worth while to keep complete records of the coal consumption, repair times and costs, the attention received in the running sheds, and so on, for each of over 10,000 locomotives separately. The task was tremendous, but the results were held to more than justify the considerable expense involved.

A fairly recent development of the application of statistics is to the control of the quality of manufactured articles. Much of modern industry is run on the lines of mass production, and this involves making separately the standard parts of an article and then assembling them. If all the parts were exactly alike, they would fit together exactly, the finished articles would be exactly alike in character and quality, and all would be well. But this does not happen. The raw materials vary in quality, important processing conditions such as atmospheric humidity and temperature vary, tools and machines are used in various states of wear, and the operators, being human, cannot work with perfect precision. The consequence of all this is that the products vary. Some components differ in size or shape from the standard so much that they will not fit in the final assembly, some are too low in quality to give a satisfactory performance. Hence arise a number of questions: What variation in quality of raw materials shall be allowed before a complaint is made to the suppliers? Must every component be inspected or only a sample? And if a sample, how big should it be and how should it be taken? How much variation in quality may be regarded as normal and at what stage does an increase in variation suggest that something has gone wrong? What is the relative importance of different factors in causing variation?

All these questions require for their answer a mixture of technical and statistical knowledge. The role of statistics is the making of systematic records of quality, the development of measures of variation, the working out of the effects of changes in quality and variation, and the application of the theory of random sampling. These problems give scope for the most advanced statistical methods available, although some simple standardized methods have been worked out for limited application.

There are published examples of this kind of use of statistics in the electric-lamp-making industry in Germany and England, in a variety of industries connected with the Bell Telephone Company in the U.S.A., and in the textile, glass-making, and engineering industries in England. The movement is spreading rapidly, and it is likely that statistics will one day be as widely used in technical control in industry as it now is in commercial control and management.

Statistics also has a part to play in the development of policy. Its first part is that of calling attention to and describing the nature of economic and social problems.

In the economic field: Is unemployment increasing or decreasing? Is it widespread or largely confined to certain areas? Which industries are expanding and which contracting? What changes are taking place in the localization of industry?

In the social field: Is there a shortage of houses? Do poverty and malnutrition abound? What changes are taking place in sickness and mortality? Is the rate of deaths due to tuberculosis higher in some areas or industries than in others? Are crime and drunkenness increasing or decreasing?

In the field of business and manufacture: Are the sales of this or that article decreasing? Are the costs of distribution in a certain area unduly high? In the factory, is there an undue loss of production because of machines being stopped for repairs?

All these are questions that may be answered statistically, and they are usually of the kind that must be answered before political action is even considered.

A second function of statistics in the realm of policy is to measure the importance of the various problems and to place them in a proper perspective. Although most economic and social problems are essentially statistical in that they concern masses or groups of people, towns, businesses, and industries, the men who have to deal with them are of the system, and since they have intimate contact with only a part it is not easy for them to see the whole. The views of the Member of Parliament are coloured by his knowledge of his constituency, which may be an agricultural or mining area; a business man tends to view all economic problems from the standpoint of his particular industry; the rent collector in a slum area sees overcrowding as the most urgent social problem; and the comfortable inhabitant of a prosperous town thinks there is nothing much wrong with the world. Some events strike our imaginations more vividly than others, either because of their nature or because of the publicity given to them. For example, a railway accident in which fifty people are killed creates a greater impression than the fact that in Great Britain during 1938 an average of over 500 people were killed in road accidents every month. In all these circumstances, the individual cannot easily take an impersonal view of things, and it is there that statistical data help.

It is, however, a weakness as well as a strength of statistics that they paint a broad, impersonal picture; for some of the 'high-lights' are lost. We have seen that statistical descriptions are essentially summaries that leave something out; and in connexion with social problems, in particular, that something is often the human touch which fires the imagination and spurs the will to action. Statistics can show the prevalence of poverty, but they cannot help the rich man to imagine what the life of the poor man is like. They can measure many of the conditions of life that promote happiness or misery, but they cannot measure happiness. The situation is well stated in the following passage from the periodical *Planning*:

'Public opinion upon many social and economic problems still suffers from an incapacity to grasp statistics, and thus fails to measure either the size of each problem as a whole or the relative importance of its different aspects. In the case of unemployment and man-power, however, the reverse appears to be true; everyone is only too ready to think in broad quantitative terms, even at the price of forgetting that rows of classified statistics are no more than feeble symbols for a multitude of men and women, each of whom individually represents a special and unique problem, which cannot be satisfactorily treated while it is simply lumped into some immense aggregate.'

The statistical description can be improved by analysing the 'immense aggregate' into sub-aggregates, as we have seen in the earlier chapters of this book, but at the best it remains only a partial description.

It is partly for this reason that the statistical investigator and the man who develops policy should not rely entirely on figures but should have direct contact with the problems with which he is dealing. The following passage by Professor A. L. Bowley suggests a realistic approach to a social problem, with a good balance between the use of statistics, particular description, and direct contact:

'If, for example, we know from the census account that in five per cent of the houses of a town there are more than two people to a room, if we ascertain that the worse houses are insanitary and small, and if we visit a few to find out the actual accommodation, the age and sex of the inhabitants and their occupations, we have probably all the data we need for criticizing or suggesting a policy of reform, without measuring the rooms or making a house to house visitation.'

I have insisted at some length on the limitations of statistics, hoping to forestall the criticisms of those who doubt the value of the subject and to temper the enthusiasms of those who over-estimate its power, but even as a dynamic of economic and social reform exact statistical description has value. The following statement made by Sir John Orr in 1936, dry and factual though it is, should have the effect of an electric shock on a man with a social conscience and an imagination:

'The tentative conclusion reached, is that a diet completely adequate for health, according to modern standards, is reached at an income-level above that of 50 per cent of the population.'

As another example of the power of statistical description, we have the following statement by Beatrice Webb, showing what effect the publication of the results of Charles Booth's social survey of London had:

'The authoritative demonstration . . . that as many as thirty per cent of the inhabitants of the richest as well as the largest city in the world lived actually at or beneath the level of bare subsistence—came as a shock to the governing class.'

Action taken as a result of an intellectual conviction derived from hard facts is likely to be more resolute than action stimulated by an emotional appeal alone; and statistical facts are of the hardest metal.

A third part for which statistics has been cast is that of acting as a guide to policy, i.e. of pointing out solutions to the problems described; but the subject does not play this part as well as it does the others. Occasionally a simple analysis of figures leads to a valuable conclusion. For example, a housing survey of Liverpool made in 1930 by the Merseyside Social Survey showed that overcrowding was not entirely due to poverty, since many a house was overcrowded even though it was occupied by more than one wage-earner earning good wages. This information at least suggests that a policy of subsidizing rents would not have solved the problem of overcrowding in Liverpool. On the other hand, there are many important questions of public policy on which neither statistics nor any other subject has given a clear answer. Is a policy of embarking on public works the best way of curing unemployment? Is an expansionist policy the best way out of a trade depression? Is a protective tariff really good for industry as a whole? How can we reduce road accidents? These are a few questions that have been the subject of opinion and controversy. The answers to them require a deep knowledge of the fundamental causes operating, but I doubt if statistics has got us very far. I discuss the place of statistics in the discovery of economic laws in the next chapter.

I have dealt in a general way with the role of statistics in the development of policy, and am now going to consider some particular aspects of the question.

The first aspect is that of economic and business forecasting. All policies are necessarily worked out for future application, and so forecasts must willy-nilly be made of future conditions. Local authorities when building schools, reservoirs, and so on must forecast requirements for years ahead; a manufacturer building a new factory will determine its size partly on his estimate of future demand for his products; most goods for consumption are ordered and made months before they are sold; and the insurance company quoting for an endowment life policy to mature twenty-five years hence must do so on the basis of an estimate of the future rates of interest and mortality experience.

The prediction of the future from a knowledge of what has happened in the past involves the belief that things will in some way continue to be as they have been. Sometimes we are naïve enough to think that the superficial appearance of things will continue to be—it is so easy to take the short view—but economic and business forecasting is based on a belief that the relations between past events have been determined by fundamental principles that are stable and will continue to be so. An empirical forecast based on a superficial analysis of past events is not very reliable, but the more successful we are in discovering the fundamental relations between events—the unchanging principles that govern change—the more scientific and reliable are our forecasts likely to be.

As an example, let us consider the prediction of the future population, which we do not expect to remain unaltered at its present level. To make a prediction we might plot a graph showing changes during the past few years, and might extend it forwards, continuing, say, the trend. In the absence of any reason for supposing that the trend will continue, that would provide an empirical forecast which might be roughly correct only for a year or two ahead. Alternatively, knowing the age composition of the present population, it is possible by assuming future birth- and death-rates to calculate the future population. Birth- and death-rates are much more stable than the size of the population, and their values for a few years in the future can be predicted, not exactly, but pretty well. Moreover, the calculation is exact, so that the resulting forecast of the population is much more accurate than the empirical one, except for the effects of factors that have not been brought into the calculation, such as migration and wars. For Great Britain, these neglected factors have not lately been very important, and much useful work has been done in predicting the future population of the country, and its age composition. However, predictions of population are probably the most scientific of any that statisticians make.

Much work has been done, too, in the development of methods for forecasting the volume of business and trade—more in the U.S.A. than in Great Britain—and these have been on various levels between the scientific and the empirical. Population is, of course, a big factor in determining economic activity, and fairly reliable long-term estimates of broad future trends in some things can be made for districts and countries by considering this. Sometimes it has been noticed that business activity at one time has been related to something that is measured at some previous time, and where this relation can be given some theoretical support it is not unreasonable to assume that it will continue with little or no change, and on that basis to make a prediction. For example, the value of houses for which plans have been passed by local authorities, figures for which are published quarterly in the Ministry of Labour Gazette, would be expected to be closely related to building activity for months ahead, and the statistics mentioned should be of value to manufacturers of building materials. The cause of such a relationship is fairly obvious. Systematic forecasts of general business activity made both in Great Britain and the U.S.A. are based on the observation that fluctuations in indexes of business activity are, broadly speaking, preceded to the extent of a few months by corresponding fluctuations in Stock Exchange prices. The theoretical justifications for these systems are (1) that movements in Stock Exchange prices represent the skilled judgement of professional operators of the state of business and its probable future course, (2) that many of the transactions in stocks of individual concerns are made by owners and other men with 'inside knowledge' of the prospects of particular businesses, and (3) movements of Stock Exchange prices are related to the same financial factors as those that have an important influence on business. The claim that the systems work is rather spoilt by the qualification that the forecasts have not been correct in abnormal circumstances.

Attempts to predict prices have also been based on the well-known theory that prices are dependent upon supply, among other things. To be able to measure the relation between supply and price for any commodity it is necessary (1) that both quantities should, for some time, have changed enough, (2) that supply should have been an important factor in causing changes in price compared with other factors, (3) that the relation between supply and price should have been fairly stable, and (4) [which is almost the same as (3)] that the conditions of supply (e.g. methods of manufacture) and of demand (e.g. the tastes or habits of consumers) should not have changed much. Articles like motor-cars and wireless sets obviously do not satisfy these conditions, but a number of primary products do to some degree, and moderately successful formulae have been obtained for predicting the prices for a few months ahead from a knowledge of the existing supply (e.g. crop) of commodities like cereals, cotton, and meat. Other factors affecting price have also been considered in the same way.

The man of affairs often needs to know how the demand for an article is related to its price—the elasticity of demand—before deciding on a price policy. One British Chancellor of the Exchequer increased the tax on sparkling wines without knowing what effect the increased price would have on consumption, and the reduction in consumption that occurred was so great that his estimate of yield was completely falsified. In 1933 the British railways took the very bold step of reducing most passenger fares from 1½*d.* to 1*d.* per mile with a view to arresting the decline in revenue. Lacking knowledge of the elasticity of demand, that important decision had to be made in the hope that increased mileage demanded by the public would be more than the one-third by which fares were reduced; otherwise revenue would have decreased. This hope was, in the event, justified. The problem of estimating the elasticity of demand is exactly similar, statistically, to that of determining the relation between price and supply. The same conditions are necessary for its solution, and similar degrees of success have been attained for various kinds of commodities.

All these problems of forecasting require for their solution economic and statistical analyses of the highest order. Economic analysis is necessary to suggest general relations that may be sought between various movements, and to explain relations that have been discovered so that they can be formulated with precision. The relations are essentially statistical in character, i.e. they are correlations, and statistical analysis is necessary to give them numerical expression. This analysis gives scope for the application of a wide range of advanced methods for separating out trends, and cyclical, seasonal, and random movements, and for disentangling the effects of the many factors involved.

I have no direct experience of business and economic forecasting, but the impression gained from reading is that its achievements are useful but modest. The forecasts are frequently fairly near to the actual events, but, frequently also, 'abnormal' conditions intervene to upset them. Business men do not rely heavily on statistical forecasts, but find it useful to consider them together with other information that can only be appraised subjectively. The word *scientific* is bandied about a good deal in connexion with business forecasts. Knowledge of the fundamental causes of economic changes is too meagre and inexact to form the basis of methods that a physicist or chemist, say, would regard as scientific, and the predictions are too inaccurate to be really scientific; but the work is proceeding along scientific lines, and as a great deal is being done we may expect that the methods will improve.

Second, I wish to consider the surveys that are now made of consumer markets and public opinion. One of the chief functions of industry is to provide the people with the things they need and desire, and there is a growing tendency among commercial concerns to embark on 'consumer research' in order to discover what the needs and desires of the people are—an activity which seems to be a department of advertising.

A little of this work takes the form of indirect statistical investigation to discover what factors influence consumers' demands. Mr. Mordecai Ezekiel in his book *Methods of Correlation Analysis* mentions some investigations made in the U.S.A. into the relation between the prices received for various products and their qualities. For example, it was found that on the Boston (Massachusetts) market for asparagus, 38½ cents extra per dozen bunches was received per extra inch of green in the stalk, 4 cents less per dozen bunches of given weight was received for every additional stick in the bunch (i.e. there was a preference for few thick sticks to many thin ones), and bunches with sticks of uniform thickness fetched higher prices than those with variable sticks. It is stated that the results 'have had a marked influence on the practices by the producers who supply the Boston market, and have led to further experimental investigation as to how to produce asparagus with desirable qualities.' Statistical investigations of these kinds are essentially applications of the methods of correlation.

Most consumer research, however, uses only elementary statistical methods. For example, localities may be compared for population and the consumption per head of (say) soap, as a preparation for a sales campaign in 'backward' areas; or districts may be compared for population and spending capacity as assessed from the occupations followed, with a view to discovering where it would be most profitable to 'push' the sales of refrigerators, washing-machines, or wireless sets.

In this field, sample surveys are much used. Before introducing a new brand of chocolates, one firm got a panel of girls to try chocolates with variously flavoured centres, and included the favourite flavours in the brand; and the preference of the public for the short-headed tooth-brush was discovered by a special sample inquiry that was made before introducing a new brand of that article. One large stores has investigated the habits of a sample of its customers to discover which ones buy at the stores regularly, in which departments they buy, the kinds of goods ordered by telephone and so on, the aim being to provide data on which to base sales and advertising policies.

Sample surveys are also being increasingly applied to the measurement of public opinion on political and similar questions. For some years American journals have conducted 'straw votes' on public questions, and the results of the Gallup Poll were much regarded and quoted when the people of Great Britain were anxiously following the development of the American attitude to the War during 1941. In Great Britain we have the corresponding polls of the British Institute of Public Opinion, the activities of Mass Observation, and the investigations of the B.B.C. Listeners' Research Department and, more recently, of the Ministry of Information. Such surveys can be reasonably reliable and can be conducted at a cost that is not prohibitive for a community. Will the development of this ability to test public opinion frequently and accurately have an effect on the theory and practice of government?

Finally, I can do no more than mention insurance, a branch of commercial activity which, particularly in the life department, is only made possible by statistics.

CHAPTER XI

STATISTICS AND OTHER SCIENCES

In this chapter I consider the relations between statistics and other branches of knowledge and take first the science of economics. There are three reasons why this subject, so regarded, would appear to be closely dependent on statistics.

First, economic laws, if they exist, refer to mass or group phenomena. Economic events are the result of actions based on the preferences, desires, and reactions of millions of people. Individually, people behave in a way that is unpredictable—some would say in a way that is indeterminate, i.e. that people have 'free-will'—

and if there are any regularities in their behaviour they are only shown in the behaviour of the mass, as described in Chapter VIII.

It does not necessarily follow, of course, that because individual people bear superficial resemblances to statistical individuals, the mass must show statistical regularities in behaviour; but such regularities do in fact seem to exist. The laws of supply and demand, for example, apply very widely; and even in time of war, when the enforcement of price regulations is backed by appeals to patriotic sentiment, 'black markets' spring into being. The belief that statistical laws do in fact describe human behaviour is implicit in the very existence of sciences like economics (and psychology), and in a rational approach to all business and political problems.

This belief does not necessarily carry with it belief in the permanence or universality of economic laws. For example, the reactions of men to financial incentives are conditioned by their ideology, which may change with time and place—the ideology of a group of Oriental mystics is very different from that of a group of English business men. Nevertheless, we believe, and act as though we believe, that most economic events follow laws which are sufficiently stable and widely applicable to be useful.

A second reason for expecting economic science to be dependent on statistics (using the word as referring to numerical data) is that the scientific way of discovering laws involves studying what actually happens, and unless the knowledge so gained is quantitative, i.e. statistical, it is 'of a meagre and unsatisfactory kind'. The immense progress that has been made in the older and more exact sciences of physics and chemistry has depended very much on measurement, and, by analogy, one would expect the same condition of progress to apply to economics.

Thirdly, if laws are to be inferred from numerical data, this must be done by methods that are largely statistical, as opposed to experimental. Economic experiments purely for the sake of gaining knowledge are not allowed; and, even if they were, it would not often be possible to isolate a few factors for investigation as the experimentalist can in his laboratory. The consequence is that the economist can only learn by observing the events that happen outside his control.

Statistical data may be used in economic inquiries in three ways: (1) they may give the information that suggests and leads to the formulation of theories, (2) they may be used for testing theories, and (3) they may provide measures of quantities that emerge from economic analysis.

(1) Economic changes can only be described by statistics. Yet the analysis of statistics does not seem to have been very fruitful in suggesting economic theories. Jevons's famous suggestion that cyclical changes in prices are correlated with sunspots was based largely on an examination of data, but it did not have a very lasting effect on economic theory. On the other hand, statistics showed the existence of the trade cycle, which has been the subject of so much economic theorizing.

(2) Economic theories have, from time to time, been put to the test by reference to statistics, but the results have not been very impressive. The whole economic system is so complicated that it is usually possible to suggest a number of theories to fit a given set of statistical facts. Consequently, when theoretical pre-

dictions have been compared with experience and there have been discrepancies, the tendency has been, not to abandon or modify the theory, but to find special reasons why the particular facts did not conform to the theory. The following quotation from some remarks made by Professor von Hayek illustrates the situation from the standpoint of one economist:

'He [Thomas Tooke] showed—and there can be little doubt about the fact—that low rates of interest usually coincide with falling prices, and high interest rates with rising prices, and concluded from this that the Ricardian idea, that a reduction in the rate of interest would lead to a rise of prices and vice versa, was wrong. I doubt whether there is to-day a single economist of repute who would be willing to assert for this reason, with Tooke, that a low interest rate leads to a fall in prices or the contrary. I need hardly waste time to explain the paradox—but statistical research has not helped us in any way to solve the difficulty.'

Tooke was presumably mistaking a correlation arising from movements in the trade cycle for a causal relationship. A modern statistician could correct for the effect of the trade cycle and obtain a partial correlation between prices and the rate of interest; but could he then be reasonably sure he had a true measure of a causal relationship? Only if theoretical analysis supported that view. Statistical analysis can separate the effects of various factors, given sufficient data, but usually only after the factors are stated by theory. Thus it seems that, in existing circumstances, theory inevitably controls the analysis of observational data and is almost unaffected by the results. It is not altogether unreasonable for economists to cling to their theories in spite of discordant statistical facts which so often are only *apparently* discordant.

In situations similar to this it is often profitable to take the theory for granted and to use the statistics to measure the importance in certain circumstances of the factors postulated in the theory. On this view, one would not have regarded Tooke's results as disproving the Ricardian idea altogether, but rather as showing that some other factor was having a much more important effect in causing the particular variations observed. The value of this viewpoint is greatest when the theoretical effects are important, but not all-important.

(3) Economic theory postulates a number of quantities such as the elasticity of supply and demand which appear only as algebraic symbols, but need to be evaluated if they are to be used. Such evaluation is a proper function of statistical methods working on statistical data. I have already, in the previous chapter (p. 158), discussed this problem as far as supply and demand are concerned, and it is only necessary here to add that although the subject is a difficult one, much work is being done in it, and progress is being made.

Altogether, the dependence of economic science on statistics and the connexion between the two subjects has not been as close as might have been expected. Economic theory has been developed, until recently at least, with scarcely any appeal to statistics for verification, and theoretical economists and statisticians have worked with comparatively little contact. Articles in economic journals and economic books contain certain algebraic formulae with unevaluated constants and formal diagrams, but very little in the way of observational data. Most statistical articles and books in the economic field, on the other hand, contain masses of data and statistical analysis, but little economic analysis.

Economists have been criticized for their alleged neglect of statistics and fact, and Professor L. Hogben has even likened them to medieval schoolmen, spinning their theories without any regard to facts. Lord Keynes wrote in 1933:

'In economic discussions Ricardo was the abstract and *a priori* theorist, Malthus the inductive and intuitive investigator who hated to stray too far from what he could test by reference to the facts and his own intuitions. . . .

'One cannot rise from a perusal of this correspondence [i.e. between Malthus and Ricardo] without a feeling that the almost total obliteration of Malthus's line of approach and the complete domination of Ricardo's for a period of a hundred years has been a disaster to the progress of economics.'

The patient collection and analysis of statistical data requires a different kind of temperament from that required for the development of economic theory, and that may be one reason why the two subjects have not come more closely together. However, it would be a mistake to suppose that economic theory is completely out of touch with reality. Lord Keynes, for example, refers to 'the amalgam of logic and intuition and the wide knowledge of facts, most of which are not precise, which is required for economic interpretation in its highest form'.

The mind can deal with facts that are not precise, whereas the more formal methods of statistics cannot; and the volume of unprecise facts is enormous compared with that of precise facts. That may be one reason why qualitative economic analysis has gone so far in spite of its comparative neglect of statistics. Another reason is probably that the economist is a man studying the behaviour of men; he sees the economic system from the inside. We may admire the high intellectual quality and penetrating power of qualitative economic analysis, and acknowledge the success it has achieved, and I would not care to say that economists have been wrong to have relied more on it than on deductions from statistics.

Nevertheless, there are limits to what can be achieved by qualitative analysis alone. Economic laws still strike the worker trained in the 'exact' sciences as being very inadequate. I think that from the standpoint of method, the analogy between economics and (say) physics is good, and economics must develop along the same path as physics by becoming more quantitative, and indeed this development is taking place. Much statistical work is being done in the economic field: work that is more than mere collection and presentation of data. The difficulties of statistical investigations in the economic field are realized and being overcome, and statistical methods are improving in power and flexibility, and are becoming more discriminative. The volume and cogency of statistical data are increasing, and particularly valuable material is being provided by the records of events and experiments like the 1931 depression, the introduction of protective tariffs in Great Britain, and the New Deal projects in the U.S.A.

It may one day happen that economics departments at universities, instead of being dominated by the theorists, will come under the domination of the statistical laboratory, just as physics and chemistry departments are dominated by the experimental laboratory.

The connexion of statistics with biology is almost as close as with economics. As long as biologists were concerned with merely describing organisms and their functions, and with classifying them into types, statistics did not come into the picture. When measurements began to be made, and the existence of variation to be recognized, statistical ideas and methods became necessary and many modern statistical methods were first developed for biological applications.

It was under the influence of Darwin's ideas and work that Galton started his numerical studies of biological variation and founded the Biometric Laboratory already referred to (p. 58). The main work of the biometricians during the early years of this century was the study of heredity in man, and of factors responsible for the 'deterioration' (as the trend was pessimistically characterized) of the race; but the scope inevitably broadened to include a variety of biological problems. This statistical approach to biology and statistical methods are twins, born and cradled together in the Biometric Laboratory.

The description of populations of biological individuals is statistical. The pages of biometric publications abound in frequency distributions of the characters of men (height, weight, reaction times to sight and sound, and so on), of plants and flowers, and of animals; and there are association and correlation tables showing the statistical relationships between two or more characters, such as the weight and vital capacity of men, the numbers of pistils and stamens in flowers, and the heights of fathers and sons. The idea behind the phrase 'like father like son' is more exactly expressed as far as height is concerned by the statement that the correlation coefficient between the heights of fathers and sons is about $+0{\cdot}5$.

The methods of statistics find their use in all branches of biology, including its applications. Applied psychologists, for example, emphasize the existence of 'individual differences' between people, but they are really calling attention to variation which they treat on ordinary statistical lines. A good deal of applied psychology is concerned with developing tests for intelligence, for skill of various kinds, for accident proneness, and so on; and the criterion of the value of such tests is that their results shall correspond with the performance of the individual in school or in some job. Such correspondence is never exact, however, so the problem of measuring its degree becomes one of statistical correlation. In this direction, psychologists have developed from the orthodox methods some variants of their own, to suit their special requirements.

Genetics is essentially a statistical subject, being concerned with the relations between the characters of groups of individuals in successive generations. The earlier work of the biometricians in this connexion was largely descriptive and empirical. This was perhaps necessary and inevitable in the early stages of the subject, and the knowledge gained has presumably been of value; but the work seems to have been almost sterile as far as the progress of genetics is concerned. A more fruitful line of attack has been based on theories developed from Mendel's discoveries. This has proceeded along statistical lines, and elaborate and highly developed statistical and mathematical methods now form the basis of a large and important branch of the subject.

Statistics is of importance in medicine. Vital statistics, epidemiology and public health, are rightly regarded as being statistical, since they are concerned with masses of people, and the treatment of these subjects is, on the whole, adequate from the statistical point of view. Statistical data and methods are also used in research in many branches of medicine—often competently, sometimes not so. Generally, there is room for increased use of modern statistical methods in medical research, and I think that even the general practitioner would be better if he was trained to be more 'statistically minded'.

In agriculture, correlation methods have been used to determine, from observations on farms not under control, what factors influence such things as the quantity and quality of crops. Such investigations have included the measurement of the effects of rainfall, sunshine, and temperature on the yield of various crops; of the relation between variations from farm to farm of the gain in weight of cattle and the quantity and quality of food; and of similar relations between the fertilizers used and the yield of various crops. This kind of investigation seems to have been more characteristic of American than of English agricultural research. In England, energy has been concentrated more on experimental investigations.

Quantitative biological experiments in all branches involve working with variable material and they provide an enormous field for the application of the sampling and experimental methods described in Chapter VII. There are also many routine biological tests of counts of bacteria in milk, of the germination of plant seeds, counts of blood particles, and so on; and the biological value of batches of substances like insulin is tested by measuring their effects on animals (e.g. rabbits) that differ in their individual reactions. In all these instances, two important questions have to be asked: 'What is the most economical way of arranging the tests to give an average result of required accuracy?' and 'How many tests must be made to attain this accuracy?' Statistical methods provide the answers, and it is only in so far as the second of these questions has been properly answered that statements of, say, the vitamin content of various preparations can be relied upon, and substances like insulin can be reliably standardized.

A great revolution took place when statistical ideas were imported into physics and chemistry—particularly the former. Physics had always been regarded as dealing with invariable constants of nature, perfectly determinate and measured with great precision. Very little room for statistics there! The conception of matter as an aggregate of elementary particles—atoms—is an old one, and contained nothing statistical, since all the atoms were alike. When, however, the particles were given different characteristics the aggregate became a statistical population and the laws of its behaviour statistical laws. This happened first with the kinetic theory of gases, in which the molecules of a gas moved in different directions with different velocities, and now applies to the modern theories of matter in terms of electrons and the like. Indeed, it seems that the only way of visualizing the electron nowadays is as a kind of a blur of probabilities. The statistical approach is not often necessary if the physicist is interested only in the properties of matter in the mass, i.e. the behaviour of the aggregate of elementary particles, but it is when he attempts to relate such properties to observations on the elementary components.

Physicists have developed their own statistical methods almost independently of the work of statisticians in other fields—there is little in common between statistical mechanics and the kinds of statistics described in this book. It may be, however, that the two branches of the subject will be related one day. Certainly the ideas used are not unique to physics.

Statistical ideas have also some place in modern chemical theories. Chemists are attempting to explain the behaviour of substances like cellulose, rubber, and proteins by postulating aggregates or chains of molecules of different lengths and weights. Changes in the behaviour of the substance as a result of chemical change are explained in terms of changes in the frequency distribution of chain length.

In both physics and chemistry, measurements made in the laboratory are subject to experimental errors. Statistical methods are somewhat used for dealing with these, although, as I have stated in Chapter VII, this application is limited. Physicists and chemists increasingly have to make measurements on variable material, however, particularly since the development of biophysics and biochemistry and the extended use of physics and chemistry in industry. Moreover, it may sometimes be necessary in technical research to make investigations or test laboratory conclusions in factories. For a variety of reasons, perfectly controlled experiments are not possible in a factory, but some degree of experimental control can often be achieved without unduly upsetting the factory routine, and a statistical experiment can be arranged. Also, a statistical analysis of the records of physical and chemical tests on output and quality, that are often kept in factories as a routine, may sometimes suggest the existence of technical effects or the causes of unwanted variations. All these situations open a wide field for the application of statistical methods to sampling, arranging experiments, and analysing and testing the significance of results.

Meteorology is usually regarded as a physical subject, presumably because the causes of weather changes are physical, but the subject also has statistical characteristics. The meteorologist has no control over weather variations, and can only record and reduce them in much the same way that we do other statistical data, using frequency distributions, averages, correlations, and so on. Many of the meteorologist's diagrams are of a special character, however, since he may require to represent at the same time, say, the wind strength and direction at various stations measured at various times of the year. Weather forecasting is based on the same logical and mathematical principles and methods as business forecasting.

In engineering experience there exists that uncontrollable variation which always indicates a field for the application of statistics. The materials the engineer uses, both raw and manufactured, vary in strength, size, and quality; loads borne by his structures and machines vary (e.g. wind pressures, the amount of traffic on a bridge, the demand for electricity); and he is often unable to control all the working conditions, such as temperature or humidity, of the processes in his charge. Sometimes, the variation is small absolutely, but must be considered because it is large compared with the precision that is required. Consequently, use must be made of frequency distributions, averages, measures of variation, and so on. For example, the strength of metal specimens is related to their hardness, and since the strength test is destructive and the hardness test is not, this relationship is of value. It is not exact, however, but is a statistical correlation, and should be regarded and treated as such. Also, sampling problems are raised in many engineering tests of materials and articles.

Engineers have their own ways of taking account of variation, but they are not always the best ways. To allow for variations in materials and loads, engineers use a factor of safety when designing a machine or structure, making the parts several times as strong as they would need to be if they all had the average strength and had always to bear the average load. These safety factors are empirical—they have been referred to as factors of ignorance—whereas it is theoretically possible to calculate them from the statistics of the variations. Such calculations have difficulties, but developments in this direction should be possible, and would almost certainly be profitable.

Another engineering way of treating variation that is not always adequate is by the use of tolerance limits. Articles delivered to a specification are not expected to be all exactly alike, but are accepted if, and only if, they are within certain tolerance limits of the specification. Tolerance limits are suitable in specifications for operations like machining, where it is easy, with care, to keep within such limits as are technically desirable, or where every article in each batch is inspected and those outside the limits can be separated and rejected. But if the inspection is by sample, the rigid use of tolerance limits involves the rejection of a whole batch if even one article in the sample is outside them, although the very fact that a sample is used implies the possibility of accepting a batch in which some articles are outside the limits but none of them happened to come in the sample. It is illogical to be willing to run this risk and yet to reject another batch because one or two articles in its sample happen to come outside the limits. Such a rejection may also be uneconomic, or it may lead to tolerance limits that are too wide to be of value. For example, limits for the life of electric lamps would have to be set at (1) a little above zero and (2) 3,400 hours, if the batch represented by the sample of Table 5 (p. 41) is to be accepted! And a batch in which all lamps had lives between, say, 200 and 500 hours would satisfy a specification containing such limits equally with a batch with lives between, say, 200 and 3,400 hours. It is important to specify not only the allowable limits of variation, but also the proportions of articles in the different regions between those limits.

The movement for the application of statistics in engineering is part of the general movement for applying the subject to technical control and research in industry. Engineers have not generally regarded themselves as needing statistical methods, but in recent years an increasing number have realized how useful the methods may be, and have applied them.

Statistics finds occasional application to many subjects. In literature, for example, frequency distributions of the lengths of sentences have been used to characterize one aspect of the style of authors. A striking and most interesting statistical investigation in the literary sphere is a study of Shakespeare made by the late Professor Caroline F. E. Spurgeon and described in her book *Shakespeare's Imagery*. She presents tables of the frequencies of various types of images used by Shakespeare in five of his plays and in certain writings of Bacon and other contemporaries. In the preface to her book, Professor Spurgeon writes:

'Shakespeare's images have, of course, constantly been picked out and drawn upon, to illustrate one aspect or another of the poet's thought or mind, but the novelty of the procedure I am describing is that *all* his images are assembled, sorted, and examined on a systematic basis.'

She also asserts that:

' . . . in the case of a poet . . . it is chiefly through his images that he, to some extent unconsciously, "gives himself away".'

Professor Spurgeon reaches one conclusion, among others: that there are two minds behind the works of Shakespeare and Bacon.

Statistical methods are also used in examining the results of psychical experiments. Such experiments are usually so arranged that their results cannot be explained by what we regard as natural causes, but require a psychical explanation if they are not attributable to chance. For example, a pack of cards may be shuffled and then turned up one at a time by an operator; a subject who cannot see the cards states to what suit each one belongs; if he is right he scores *one* and if not he scores *nothing*. The question then arises: is the subject's score greater than can be attributed to chance, i.e.? is it greater than it would have been had he guessed the cards? The answering of such a question involves a regular use of standard statistical methods.

For some subjects statistics provides ideas of basic importance; for some it provides methods of investigation. In one way or the other, or in both ways, statistics has an impact on most other branches of knowledge. In this respect it is not unlike arithmetic. Arithmetic is so woven into the fabric of our thinking that we use it almost subconsciously, and, after leaving school, most of us are scarcely conscious of its existence as a separate department of study. On the other hand, most people are scarcely conscious of statistics except as a separate subject. I look forward to the day when statistics will occupy a place in education only a little way behind arithmetic; when everyone will learn as much of the subject as is necessary for ordinary life and for his particular vocation. Then everyone will use statistics easily and naturally, and such general introductory books as this will become obsolete.

NOTES ON BOOKS

EXCEPT for the first, the books in the following list are more for the serious student than for the general reader. Nevertheless, anyone interested in the various subjects can usually learn something from reading parts of the books, even if he does not apply himself to studying the whole.

My Apprenticeship by Beatrice Webb gives an interesting first-hand account of the life and work of a social investigator. *Methods of Social Study* by Sidney and Beatrice Webb and *The Measurement of Social Phenomena* by A. L. Bowley are more for the serious worker but are quite easy to read.

Management, Planning, and Control by A. G. H. Dent shows the uses of statistics in scientific management and gives a good bibliography. *Business Forecasting and its Practical Application* by William Wallace gives the attitude of one who has first-hand contacts with both statistics and business.

Statistical Method in Economic and Political Science by P. Sargant Florence gives a very full, general discussion of the relations between the subjects in its title.

Workers in the social, business, and economic fields have available a very large number of textbooks on statistical methods, of which *Elementary Statistical Methods* by E. C. Rhodes is a good one to start with, and *Elements of Statistics* by A. L. Bowley is a good one to follow with.

Experimentalists, especially those working in the biological sciences, will find *Statistical Methods* by

George W. Snedecor a good introduction as well as a textbook. *Statistical Methods for Research Workers* and *Design of Experiments* by R. A. Fisher are important books, but the reader will not go far in them until he has learnt something of statistical methods.

Medical readers will find *Principles of Medical Statistics* by A. Bradford Hill a good introduction and *An Introduction to Medical Statistics* by H. M. Woods and W. T. Russell a first textbook.

Application of Statistical Methods to Industrial Standardization and Quality Control by E. S. Pearson and *Quality Control Charts* by B. P. Dudding and W. J. Jennett deal shortly with general principles and give simple procedures for industrial application. *Economic Control of Quality of Manufactured Products* by W. A. Shewhart is a fuller treatment of the whole subject. As a textbook there is *An Engineer's Manual of Statistics* by L. E. Simon.

The textbooks so far mentioned are for particular applications of statistics. The classical general textbook is *An Introduction to the Theory of Statistics* by G. Udny Yule and M. G. Kendall.

INDEX

PRINTED IN
GREAT BRITAIN
BY
THE RIVERSIDE PRESS
EDINBURGH