The Application of the Statistical Method in Industry

P. C. Mahalanobis B. A. (Cantab), I. E. S.
Presidency College, Calcutta.

Introduction.

Until recently the statistical method in industry has been associated with practically one single group of problems, namely marketing. But during the last 3 or 4 years, rapid developments have taken place in the application of the statistical method to technical problems in industry. There is a growing realization of the benefit which industry may gain by the use of analytic statistics to a whole range of problems that arise in the course of routine production such as the purchase of raw materials, control and standardization of the quality of manufactured products; problems of sampling and testing etc. There are also special research problems in industry which involve careful planning and experiment, and which require the same kind of statistical control and analysis as agricultural field experiments of the new type. An attempt has been made in the present note to give a general idea of the scope of the statistical method in industry, but owing to the highly mathematical nature of the subject it has not been found possible to touch any of the technical details.

Control of Purchase of Raw Materials.

In making bulk-purchases (whether of raw materials or of partly or fully manufactured products) it is usually not practicable to test every unit of the purchased goods. For consumable commodities like coal or electrical fuses, which are destroyed during the process of testing, such complete tests are of course even theoretically impossible.

Thus the need constantly arises for estimating the quality of a large batch or consignment of similar material from the inspection of a small amount drawn from the larger whole. We are therefore obliged to have recourse to the method of sampling tests carried out on a small fraction of the total purchases. (The term 'sample' is used in this note in its statistical as distinct from its ordinary commercial sense. Thus a 'sample' consists of a number of individual units for each of which one or more characters have been or are going to be measured and recorded; it is often usual to speak of these measured values themselves as forming the sample).

From the purchaser's point of view three things are important :—

(i) the goods purchased must attain a specified standard ;

(ii) the quality received by him shall remain uniform from one consignment to another ; and

(iii) the amount of sampling tests he has to carry out should be reduced to a minimum, consistent with safety.

We must take into consideration the existence of variation in this connexion. For example, the percentage ash-content of coal from the same source of supply may vary by 50 or 100 per cent; different test specimens from the same batch of yarn vary in strength by 50 to 100 or more per cent; the length of life of electric lamps of marked voltage may vary from one lamp to another by 200 or 300 per cent. and so on. Testing a single unit will clearly be inadequate. The question naturally arises, how many units must be tested in order to attain a given degree of accuracy? How should the units in the sample be chosen? For example, in sampling raw cotton, is one handful representative of a bale, or is one bale representative of a consignment? Is it better to choose 500 hairs singly, say 50 from each of the 10 bales in a consignment, or should they be chosen in tufts of 50? In sampling coal, should we take a handful from a single sack, or take samples from a number of sacks? In testing electric lamps, should we take 100 samples of 2, or 40 samples of 5 or 2 samples of 100 lamps? When it is remembered that there may be variations within a unit; between units in the same consignment; between consignments from one source of supply; between sources of supply; and between supplies at different seasons, the complexity of the problem will be easily realized.

We have already mentioned that the existence of variation must be accepted as a basic fact of experience. The knowledge of 'mean value' alone is therefore practically useless, and it is essential to take into consideration the nature and amount of variation present. This emphasis on the study of variation distinguishes modern analytic statistics from the older enumerative statistics in which totals or mean values only are of importance. The magnitude of variation is usually measured by what is technically known as the "variance", or its square root, the "standard deviation". The 'variance' is simply the average value of the squares of all deviations from the mean, and is obtained by taking the sum of squares of all deviations from the mean, and dividing this sum by the total number of deviations. It is clear that the greater the magnitude of variation, the larger will be the value of the variance.

Statistical control is based on the important fact that for homogeneous material both the mean value and the variance remain sensibly constant. In fact such stability of the mean value and of the nature and magnitude of the variation (as measured by the variance) are the only criteria of homogeneity. The size and number of samples will for example obviously depend on the magnitude of the variance. Other things being equal, a larger size or a larger number of samples will be required to attain the same degree of accuracy when variance is large than when variance is small. The probable limit of variation from one sample to another will also clearly depend on the value of the variance.

Thus once the mean value and the variance are known, it becomes possible to prepare a complete testing scheme on the following lines:

(1) A standard programme is prepared for routine testing which will include detailed instructions regarding (i) the size of samples, (ii) the method of taking the samples, (iii) the number of samples, and the interval at which they should be taken. It is usually convenient to arrange for the results of the sampling tests being plotted on graphical control charts.

(2) An upper and a lower "limit of safe variation" will be marked on the control chart, such that so long as the observed results of sampling tests fall within these limits it may be assumed that the quality of the material is controlled and is remaining sensibly uniform.

(3) A second pair of "danger limits" lying outside the "safe limits" will also be marked on the charts. If the result of a sampling test falls outside these danger limits it will signify an appreciable departure from uniformity which requires special investigation.

(4) If the result falls in the region lying between the safe and the danger limits, that is, falls in what may be called the "warning belt", it will indicate that the situation is suspicious. It will then be advisable to repeat the sampling test until it becomes clear whether the control has broken down or not.

The advantages of such a scheme of statistical control are many. It will reduce the cost of testing to a minimum, consistent with safety. It will give immediate warning of any suspicious increase in variation, and will give definite indications of any appreciable departure from uniformity of quality. Once the control system is properly designed it will practically work automatically until the danger zone is reached, when of course special investigations will again become necessary. Finally the method is extremely elastic, and the testing programme can be adjusted to suit any desired degree of accuracy.

The degree of accuracy itself must of course be fixed from considerations of a practical nature. If the specifications are stringent the cost of testing will be large, for larger or more numerous samples will be required. An inadequate specification on the other hand will cost very little in sampling tests, but will be practically useless. There will be an optimum degree of accuracy and amount of testing for a particular material or industry, and this optimum must be determined with the help of statistical analysis.

Routine Control of Quality of Manufactured Products.

Similar statistical problems arise in attempting to control the quality of manufactured products. Even when the quality of the raw material is kept fairly uniform, numerous factors such as temperature, humidity, rate of working etc., cannot be fully controlled, and consequently the product varies in quality. The statistical aim at this stage is two-fold. The first objective is to eliminate assignable causes of variation in quality. For example, in many industries the quality of the product varies considerably with humidity. A manufacturer may therefore want to know whether it is worth incurring the heavy expenditure of installing a plant for artificial humidification and air conditioning. By correlating daily output and quality of the manufactured product with daily humidity records for the (as yet) unhumidified factory, the statistician can often make an estimate of the probable gain in production to be expected from the installation of a plant, and thus help the manufacturer in making a decision.

In many instances the economic efficiency (or rather, inefficiency) of production is measured by the the fraction of the manufactured product which fails to conform to the engineering specifications, or by what is usually called the fraction defective. Under controlled conditions this fraction should fluctuate within a normal range of variation. If it is found that the variation is irregular and often greater than this normal range, we shall have every reason to suspect the presence of some gross or definite cause of variation. Now it is clear that even when the average quality of the product remains constant, the proportion of defectives will inevitably rise when the range of variation increases. To reduce and control the magnitude of variation is thus of prime importance in such cases. All assignable or systematic causes of variation (the existence of which is indicated by statistical analysis) must therefore be eliminated by suitable improvements or changes in manufacturing processes until the statistical test assures us that the fluctuation has been brought within normal limits.

Once the assignable causes of variation have been eliminated, it becomes possible to attain the second object of statistical control, namely, the setting up of a routine programme of testing which should be adequate but not too costly. The plan for such a test programme will naturally be very similar to that already described in paragraph 3, and will consist of detailed instructions regarding size and number of samples, methods of collecting them, and control charts with double sets of 'safe' and 'danger' limits.

The setting up of such a system in a factory, the elimination of assignable causes of variation by gradual changes and improvements in the process of manufacture, and the use and interpretation of routine tests for investigating causes of variation require an intimate acquaintance with the working of the factory as well as the most refined statistical technique. The essential need for co-operation between engineers and statisticians cannot therefore be overemphasized.

Specification and Standardization.

The growing strength of the movement for national and international standardization of materials shows the industrial importance of the subject. The degree of accuracy of specifications and of standardization which can be attained in practice ultimately depends on the variability of the quality of the manufactured products, and on the reliability with which such variability can be estimated.

It is being more and more widely recognized that uncontrolled variation is the most potent source of inefficiency in every direction. Once a fairly satisfactory level of average quality is attained, a reduction in variability becomes usually more important than a further improvement of the quality. Consider, for example, the tensile strength of malleable ironcastings. Even when the average strength is high, if variability is large, the marginal "factor of safety" must be kept large in order to ensure the strength not falling below the minimum requirement. Provided the variability can be appreciably reduced, it will be clearly possible to lower the "safety factor" without any increase in the average strength ; in fact when variability is small such reduction of the "safety factor" may even become possible with a lower average quality. It is likely therefore that many of the "safety factors" used in engineering practice could be reduced without danger, provided the variability were reduced and estimated with greater precision.

Standardization of specifications thus calls for reliable estimates of the average quality as well as of the variability in the quality of the material. The need for sampling programmes and of the use of statistical analysis for this purpose is thus clear.

Problems of Management and the Human Factor.

Statistical methods are equally indispensable for the analysis of variations caused by the human factor. In fact statistical analysis has functioned as the basic tool in practically recent investigations in industrial psychology in dealing with a large variety of problems such as the influence of temperature, humidity, monotony, rest-pauses and other conditions of working on the quality and quantity of output ; investigations on accident-proneness and prevention of industrial accidents ; selection of employees ; efficiency of methods of training ; fluctuations in labour turn-over ; time-studies of the rate of production in relation to stoppages of machines and unequal distribution of work ; comparative efficiency of operatives and methods of organization etc. Besides these recent developments, the statistical method has been, and is still being, extensively used in connexion with the study of marketing problems of all kinds including cost analysis, comparative efficiency of advertisements, market surveys, business forecasting and industrial planning.

Statistical Control in Industrial Research.

The primary object of industrial research is the comparison of different kinds of raw material, of different types of machinery, or of different technical processes to find out which gives the better or more uniform quality of product or a larger output. Under laboratory conditions the different factors can often be isolated and studied separately, but this is usually impossible under actual factory conditions of production. It therefore becomes difficult or impossible by ordinary laboratory methods to disentangle the different causes and effects in order to study the relationship between the factors under investigation. Recourse to the statistical method is thus inevitable, and it is necessary to consider statistical principles in designing the experiments. The general aim is to try to balance out all the uncontrolled factors of variation. The basic problem from the statistical point of view is practically the same as in agricultural experiments of the new type in which the effects of different "treatments" are compared with the help of suitable randomized designs. For example, Dr. R. H. Pickard D. Sc., F. R. S., Director of the British Cotton Industry Research Association, has explained in a recent paper how the well-known "Latin Square" design of agricultural field experiments was used with great success in an investigation on the influence of "wettability" of a cloth on the rate at which number of experiments could be reduced by half for given standard of significance.

It is claimed that the use of the modern statistical technique in industrial research has the following advantages :—

(1) Owing to the statistical control in the design of the experiment it is possible to obtain valid results by eliminating the effect of chance factors.

(2) The precision of the comparisons is increased appreciably, and all results are obtained in terms of the probability scale.

(3) It usually becomes possible to reduce considerably the total number of experiments thus effecting a substantial saving of both time and money.

(4) The simultaneous study of a number of factors is rendered possible by a design of a "complex" type. It thus becomes possible to study the "inter-action" or mutual reaction of different factors. Such studies are altogether impossible in the older type of research in which the different factors are investigated separately and in isolation.

(5) With the help of correlational analysis it becomes possible to study the inter-connexions between different sets of factors, and thus obtain a greater degree of control in many cases.

Advantages of Statistical Controls in Industry.

It will be seen from the preceding discussion that statistical analysis can provide industry with a scientific basis for the establishment of economic standards of quality, and an essential tool for the maintenance of such standards under routine conditions of large scale production. The use of the statistical method will give adequate control over the quality of materials purchased in the bulk; reduce the cost of routine inspection and testing; help in reducing the magnitude of fluctuations in the quality of manufactured products, and by eliminating assignable causes of variation reduce the proportion of defective; make it possible to reduce without danger the 'factors of safety'; make specifications more reliable, and thus help international standardization; furnish an effective tool for the study of problems of management involving the human factor; and finally increase the efficiency and reliability of experiments in industrial research.

Actual experience has shown that the cost of using statistical controls is comparatively a small item, and is usually fully worth while from a purely business point of view. It is claimed in fact that the statistical method is indispensable for the elimination of waste, and for attaining economic standards of efficiency in production.

Recent Developments in Western Countries.

Dr. W. A. Shewhart of the Bell Telephone Co., U. S. A. published in 1931 an important book on the *Control of Quality of Manufactured products* dealing exclusively with the application of the statistical method (especially the theory of sampling distribution) in industry. A little later a Joint Committee was set up for the development of statistical methods in the U. S. A., by the American Society for Mechanical Engineers and the American Society for testing Materials. Dr. Shewhart was invited by the University of London to give a course of lectures in London in May 1932 on the *Role of Statistical Theory in Industrial Standardization.* This immediately led to the British Standards Institution calling a conference of representatives from several different engineering groups, scientific societies and research institutes, and a committee was appointed under the chairmanship of Mr. B. H. Wilsdon of the Department of Scientific and Industrial Research to draw up a report on the subject. Similar action was taken in Germany at about the same time by the Deutscher Normensausschuss.

Dr. Egon Pearson of the Biometric Laboratory, London, contributed a paper on the quality-control of output to the York meeting of the British Association, and read an important paper before the Royal Statistical Society in December, 1932, on *Statistical Method in the Control and Standardization of the Quality of Manufactured Products* which attracted a considerable amount of attention. The subject was discussed in a special article in *Nature* on October 29, 1932, and again on December 24, 1932, and it was suggested that provision should be made in the Biometric Laboratory for the development of education and research in this subject. The question was immediately taken up, and a new Department of Applied Statistics with Dr. E. S. Pearson at its head was created by the University of London in October 1933.

In the meantime a Committee had been appointed by the Council of the Royal Statistical Society for forming an organization which would provide facilities for the consideration of the problems involved in the application of statistical methods to industrial and agricultural research: and a new 'Industrial and

Agricultural Section' of the Society was formed on November 23, 1933, when Dr. R. H. Pickard, F.R.S., Director of the British Cotton Industry Research Association, read a paper on the *Application of Statistical Methods to Production and Research to Industry.*

A Scheme for India.

In view of the practical importance of the subject it is suggested that provision should be made for its study and its application to problems of industrial production, organization and research in India without further delay. The introduction of the statistical method is likely to prove immediately useful in India in connexion with the testing of raw materials and manufactured products of all kinds, in railway workshops, telegraph and telephone engineering etc, and in such industries as textile, iron and steel, jute, coal, sugar, leather, paper etc.

It is suggested that the actual testing and the collection of primary data will be done by engineers and technologists under factory conditions or in testing laboratories, and the data will be reduced and analysed in the statistical laboratory. The statistician will prepare testing programmes, while the technologist will try them out under actual working conditions and send the results to the statistician for analysis and interpretation. In the beginning the work will necessarily be of the 'trial and error' type, but with the accumulation of data relating to the variability of the material or product under investigation, it will gradually become possible to set up a permanent programme with standardized control charts. This will enable the work of testing to proceed on routine lines without further reference to the statistician until there is a break-down of the control. Again in special investigations (rendered necessary by such break-downs, or in industrial research) when assignable causes of variation will have to be traced and eliminated, the design for conducting experiments will be prepared by the statistician, while the actual exprimental work will be done by the technologists. In this way by close co-operation between technicians and statisticians the complicated problems of industrial production can be studied in all their intricacies in the most efficient manner.

India is admittedly backward in industry. The need for using modern methods of scientific research is therefore all the more urgent. Analytic statistics has been found to be a most powerful tool for this purpose in the western countries. It is bound to prove even more useful in India where urgent problems of all kinds require solution in every direction.

The application of the statistical method has already been started in a small way, mainly on the theoretical side, in the Statistical Laboratory, Presidency College, Calcutta. An attempt is being made to keep a record of progress in various fields in western countries. Detailed information and free advice on any point will be gladly supplied to industrialists, technologists and other persons, interested in Indian industries. The Laboratory is also prepared to undertake research problems of all kinds including the preparation of control charts, designs for routine sampling and technical experiments, interpretation of statistical data, and to cooperate in other ways for advancing the cause of industrial progress in India.