Statistical Methods In

Engineering Problems

By

W. A. Shewhart
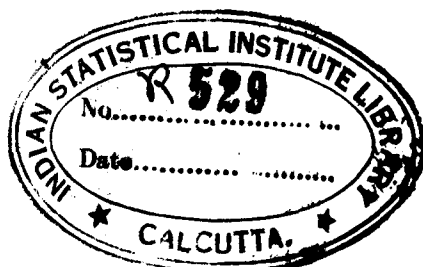
Statistical Methods in Engineering Problems

Engineering methods of calculating averages and errors of averages have not kept pace with the rapid developments in statistical theory

By W. A. Shewhart, Ph.D.*

Bell Telephone Laboratories, Inc.

ERE and there in the engineering literature of recent years we may find an occasional reference to the application of statistical methods to particular engineering problems. Furthermore the necessity for estimating the probable errors of experimentally determined results has been accepted in general by the engineering profession.

This note calls attention to some additional applications of modern mathematical statistical theory, to research, design, production, inspection, supply and other engineering problems. Attention is given to certain general types of problems in the solution of which statistical applications have been made, and to the nature of the possible economies effected thereby. It is reasonable to believe that very definite economic advantages can be obtained in any large industry through such applications.

In the application of statistics, as in the application of scientific laws in general, it is often very difficult to form even an approximate estimate of the magnitude of the resulting economic advantages and hence no effort will be made to give specific figures.

Some very important problems call for statistical solutions. One general type is the estimation of safety factors to allow for the happening of the unexpected. The failure of a piece of apparatus may involve, in addition to great financial loss, a serious injury or even death to one or more individuals. Telephone engineering illustrations could be mentioned but we shall consider a problem arising outside this field.

The civil engineer often must build levees or dams to take care of exceptional rainfall. Customarily he must make his estimate of the limiting flood height upon the basis of a few yearly records of the yearly run-off of the given area. Too low an estimate may result in a destructive flood. Of course this gives rise to a difficult problem in the solution of which modern statistical theory may prove to be a helpful tool.

Setting Engineering Standards

ANOTHER general problem is that of setting engineering standards, where, as is often the case, the underlying experimental data may show considerable dispersion. For example, the strength of a given kind of telephone pole is a function of its modulus of rupture. If we use standard methods and measure the modulus of rupture of each of a number of poles which appear to form a homogeneous group, we must expect to find even wide differences between the observed values because there are many factors influencing the modulus of rupture not easily discernible or at least not easily measurable and controllable. Indications are, however, that these causal factors approximate a constant system of chance causes, and hence that probability theory is applicable.

Assume that these differences in observed values of modulus of rupture for the different poles can be sufficiently well characterized by two parameters; viz., the average and the standard deviation. It is of great economic importance to obtain the best possible estimates of the expected values of these two factors as we shall now see.

If the average taken as standard happens to be in excess of the expected modulus, serious losses may be incurred through breaking of telephone poles under storm conditions; on the other hand, if the chosen standard happens to be below the expected value, losses estimated in hundreds of thousands of dollars are incurred through wholesale removal of telephone poles a number of years before such removal is actually necessary. Modern statistical methods assist an engineer in forming the best estimate of such standards in terms of the observed data, by making it possible for him to allow for sample size, to correct for error of measurement and to insure a homogeneous sample upon which to base his estimates.

Estimating Errors

THE results of research are almost always expressed in terms of averages and errors of averages, but engineering methods of calculating such errors have not kept pace with the rapid development in statistical theory. Particularly in cases of small numbers of measurements, errors calculated by the customary methods are much too small. In fact our best estimates may be 100 per cent or 200 per cent higher than those customarily used.

Even on general principles it goes without saying that, if it is worth while making an estimate of an error, it is worth while doing it to the best of our ability. Nothing short of the best estimates can satisfy the modern research engineer, particularly since these do not involve any additional labor. However, in certain classes of cases the economic importance of attaining the best estimate of an error can be easily illustrated. One such is the determination of the significance of observed differences in the measurements of the physical properties of two kinds of materials or in the qualities of operation of two experimental models of a given kind of apparatus. A specific instance would be that of comparing the electrical characteristics of two contact alloys, one the standard now in use and another a much cheaper material. In such instances we have two sets of experimental data corresponding to either the two kinds of material or the two kinds of model. We must determine if the observed differences between the two sets are reasonably attributable to chance. If they are and if we must select one or the

upon the basis of the available data without further experimentation, then we are free to choose the most economical material or design, even though the observed quality of the cheaper material is not quite up to standard.

Another application arises in building up any kind of apparatus out of a large number of piece-parts. To secure the economies of wholesale production, piece-parts are customarily manufactured in large quantities. It is a well-recognized fact, however, that the piece-parts of any particular kind are not identical one with the other and that the quality of operation of any assembled piece of apparatus is a function of the qualities of its piece-parts. To assure that a particular piece of apparatus will function within prescribed limits when it is built up of piece-parts selected at random, it is therefore necessary to control the probability distributions of each of the parts so that the resultant the chance fluctuation in a given part. Hence it insures an variations of the various piece-parts.

Estimating Degree of Dependence

MODERN statistical theory shows the cost of reducing the overall chance fluctuation by modifying the chance fluctuation in a given part. Hence it insures an economic distribution of effort in controlling the chance variations of the various piece-parts entering into the system.

Perhaps one of the most general and at the same time most difficult problems is to estimate the degree of dependence of a property of some material or of a piece of apparatus upon some one of its physical characteristics. For example it is known that the moisture content of telephone poles affects their modulus of rupture. Indeed a very definite relationship can be established between moisture content and modulus of rupture of small sawed pieces of timber. However, to determine such a relationship for poles is difficult because we cannot control all factors other than moisture content. The effects of these uncontrolled factors mask that of moisture content so that the degree of relationship between modulus of rupture and moisture content for telephone poles is a problem involving the theory of correlation. Examples of this character could be multiplied in all fields of engineering work.

Objects of Inspection

INSPECTION engineering has two objects: to protect the consumer and to effect economies in the method of production. To see how these two objects are attained it is of interest to consider briefly some of the details of the manufacturing problem. The telephone instrument, which is so familiar to everyone, is not so simple as it looks. To make it requires 201 parts and to connect it to another instrument requires approximately 110,000 more parts. The annual production of most of these parts runs into the millions, so that the total annual production of parts runs into the billions.

Twenty or more raw materials such as gold, platinum, silver, copper, tin, lead, wool, rubber and silk literally collected from the four corners of the earth, are used in the manufacturing process to produce this great quantity of piece-parts. The 100 per cent inspection given all telephone equipment at the time of installation attains the first object — the protection of the consumer. It is, as it were, the consumer's watch-dog. Modern applications of statistical theory to inspection, however, have to do with the second object, the economic phases of which can best be divided into five parts:

1—To determine, for each step in the process inspection, the economic percentage rejection or tolerance for defectives.

2—To reject defective material at such points in the chain of production as will make the net cost of rejection a minimum.

3—To determine for each step the minimum amount of inspection which will suffice to give economic control of quality. In other words, to introduce sampling inspection wherever possible.

4—To detect lack of control of quality or trends and erratic fluctuations.

5—To assist in finding the causes of these trends and to assist in their control wherever necessary.

Reducing Inspection Costs

THROUGH application of statistical theory, a very appreciable percentage reduction in the cost of inspection has been attained. The estimated annual savings to the Bell System accruing from carrying out the third object alone more than justifies all of the attention to such studies.

In production engineering where there are, in general, large quantities of data to be analyzed, even the choice of the method of analysis to be used may lead to appreciable savings through reduction of the number of observations required. A typical case is that where the estimate of the standard deviation may be obtained by either the root mean square or mean error method. Because of the greater efficiency of the root mean square method, fewer observations are required to obtain a certain degree of precision with this method than with the other one and annual savings effected in this manner may run into the thousands of dollars for a single kind of apparatus.

All Engineering Data a Sample

WE shall close with a supply problem. Telephone poles are supplied to the trade in a comparatively large number of classes to meet the various needs. To be specific, let N be the expected number of poles per acre, m be the number of sub-classes of poles defined in the specifications, $p_1'N$, $p_2'N$, $\ldots$, $p_m'N$ the estimated expectancies of poles per sub-class per acre, p_1M, p_2M, $\ldots$, p_mM the estimated expectancies of poles supplied to the trade per sub-class upon the basis of past experience, X_1 be the total expected cost of poles if $p_1'M$, $p_2'M$, $\ldots$, $p_m'M$ of each class are used and let X_2 be the total cost if p_1M, p_2M, $\ldots$, p_mM of each class are used, where M is the total number of poles used per year. Can we effect an appreciable annual saving by taking all possible steps to make $p_i' = p_i$ for $i = 1, 2, \ldots, m$?

Of course the two sets of estimated expectancies are subject to sampling errors and hence the observed differences between the two sets may or may not be attributable to chance. If they are, it would be useless to try to effect a savings in the way just indicated. On the other hand, if some test such as Chi Square indicates a significant difference between these two sets, then an attempt may be made to secure the economies accruing from making $p_i' = p_i$ for $i = 1, 2, \ldots, m$.

In one such study preliminary results indicate that an annual saving running into thousands of dollars may be made possible by securing a consumption of poles more nearly consistent with the expected frequency distribution into sub-classes as given by nature.

From a statistician's viewpoint, all engineering data may be considered as a sample. The engineer and man of business have for several years been making applications of statistical methods for representing these data. In fact it is this phase of the subject which up to the present time has received practically all the consideration in textbooks on the application of statistics to industry. It will be noted, however, that the problems raised in this paper are of a different nature and have to do with making use of data constituting past experience in guiding future engineering activities. All of these problems call for sampling theory and thus throw open an almost uncharted field of application. It is the interpretation of causal relationships underlying sampling theory that appeals to the engineer who himself believes, as did Thiele, that "Everything that exists and everything that happens is a consequence of a previous state of things."