SCIENCE

Vol. 97 | Friday, March 5, 1943 | No. 2514

SCIENCE: A Weekly Journal devoted to the Advancement of Science, edited by J. McKeen Cattell and published every Friday by

THE SCIENCE PRESS

Lancaster, Pennsylvania

Annual Subscription, $6.00 Single Copies, 15 Cts.

SCIENCE is the official organ of the American Association for the Advancement of Science. Information regarding membership in the Association may be secured from the office of the permanent secretary in the Smithsonian Institution Building, Washington, D. C.

OPPORTUNITIES IN MATHEMATICAL STATISTICS, WITH SPECIAL REFERENCE TO SAMPLING AND QUALITY CONTROL[1]

By Dr. W. EDWARDS DEMING

BUREAU OF THE CENSUS AND BUREAU OF THE BUDGET

Seed haunted by the sun never fails to find its way between the stones. And the pure logician, if no sun draws him forth, remains entangled in his own logic.
—*Antoine de Saint-Exupéry, The Atlantic,* March, 1942: page 328.

The control chart was devised by Shewhart in 1924 to help disclose the presence of extraneous causes of variability that are worth looking for; also to give greater quality assurance in devising acceptance procedures (Problems B and A, respectively, as outlined below). If this were a group of business men, I might seize this opportunity to persuade you to make use of these methods. But speaking before mathematicians, I need not do that. Here we can talk about the next step, *viz.,* how to harness the efforts of mathematicians to statistical problems.

I shall remind you of two problems that confront the manufacturer and the statistician in industry:

Problem A: What to do with this lot? (Accept it, reject, pass, scrap, rework, or regrade it)

Problem B: What to do with the process? (Leave it alone; or look for some identifiable cause, make some adjustment, use different raw materials)

The quality control engineer does his best work in either problem when he recognizes the existence of both, and deals with both simultaneously. In par-

[1] An address given at a joint session of the Institute of Mathematical Statistics and the American Mathematical Society, Vassar College, on September 9, 1942.

ticular, the decision whether to accept or reject a lot of manufactured articles can not be dissociated from the question of control: is this lot one of a series of lots that were made in a state of statistical control? If the answer is no, each lot must be examined in considerable detail if its quality (average diameter or per cent. defective) is to be determined accurately. If the answer is yes, and if the lots are large, inspection of a small percentage of each lot, or of every k-th lot, will give the needed quality assurance. This *increased quality assurance* with *decreased inspection* is one of the advantages of the Shewhart methods.

The Shewhart methods are equally important in the social sciences, where a similar division of problems exists, though with some necessary modifications.[2] With the increasing number of surveys of local and national scope being carried out to supply data on the social and economic states of the population, and on natural resources and stocks, and with the ever-increasing need for accuracy in predictions, the requirements being placed on sampling methods and the interpretation of data are assuming increased severity. Recognition of the importance of sampling is one of the things that brings us together here.

Statisticians and mathematicians are rendering meritorious service in these critical times. I shall try to show you how they can render still more.

Variations in Nature

He who ignores the inherent variability of all nature does not understand nature. You may leave home at the same minute morning after morning, but you do not always arrive at your office at the same time. The peas in a pod are not alike, but they are usually alike enough for the purpose. Two telephones are not alike, but the variations are within limits that are known to be close enough and known in advance of production. Traffic accidents, deplorable though they are, do happen, and what is more, *will fluctuate week by week,* even under a constant set of causes. Likewise, a sentence will have various meanings, depending on who reads it. The meaning of anything that you say is the *action that it produces.* In attempting to write clearly you can only hope to narrow the spread of the interpretations—*i.e.,* of the actions or emotions that your writing gives rise to when readers read it.

I heard that the president of one large corporation was always anxious whenever the amount of scrap was greater one day than it was the day before. Now of course it is well to get interested in scrap, and the control of scrap incidentally offers one of the best uses of the Shewhart control chart; but if one looks for trouble in the plant every time the amount of scrap is higher one day than it was the day before, he will be looking for trouble far too often. It costs money to look for trouble. The amount of scrap is going to vary day after day even under controlled circumstances (*vide infra*).

[2] These remarks are extended in my paper in the *Jour. Am. Statistical Assoc.,* June, 1942, Vol. 37, pp. 173–185.

Definition of a Constant Cause System

A constant cause system may be described as one that produces numerical results that vary from one to another like a sequence of numbers that are drawn from a bowl of physically similar numbered chips, blindfolded, and with replacement and shuffling. Drawing from a bowl is the limiting attainable state of knowledge, which is to say, this is the limiting attainable state of stability.

The prime requirement of a constant cause system (a controlled process—*vide infra*) is stability, and hence the predictability of future percentiles or probability tolerances. It is important to recognize that a constant cause system does not produce constant results.

Randomness, Statistical Control and Technical Control

Any observation, measurement, or table of frequencies, comes about as the result of applying an operation. Whatever the operation, a repeated application thereof gives a new observation. Continued repetition gives a sequence of terms—a *population* of results. Thus, suppose an article is being manufactured, one after another, and let each be measured for its outside diameter. The situation is something like this—

$$\begin{array}{ll} \text{Article:} & F_1,\ F_2,\ F_3,\ \ldots,\ F_n \\ \text{Diameter:} & X_1,\ X_2,\ X_3,\ \ldots,\ X_n \end{array}$$

The manufacture and the testing and the recording of the diameters constitute the operation of obtaining the n values of X.

In the state of randomness or stability, the terms of the sequence behave as if they were being produced by a constant cause system—they behave as if they were being drawn from the bowl. Without stability, probability tolerances for future terms can not be set purely on the basis of past terms.

The attainment of randomness is much more difficult than is commonly supposed. It may require months. Statistical control is more than technical control. In the latter, the terms of the sequence are merely kept within certain bounds. Be they ever so narrow, this is not statistical control. Of course, technical control comes first, and is often sufficient.

A practical criterion of stability is the Shewhart criterion. If 25 or more successive points, each consisting of four consecutive terms of a sequence, fall within the control limits, after it is thought that all possible sources of extraneous variability have been eliminated from the process that gives rise to the sequence, and if no trends or patterns are exhibited by the sequence, it is possible to set probability tolerances on future terms of the sequence, and thus to predict how much of tomorrow's product will fall within various limits—possible in the sense that experience, in many different kinds of production, covering many years, has borne out the advisability of so doing. I shall return to this topic later.

THE FUNDAMENTAL PROBLEM OF MATHEMATICAL STATISTICS

The fundamental problem of mathematical statistics is to set "probability tolerances," *i.e.*, to set limits within which 99 or some other per cent. of the next (*e.g.*) 1,000 or 10,000 terms of a random sequence will fall. For instance, on the basis of the past n diameters in the example cited above, the problem is to say what proportion of the next 1,000 or 10,000 diameters will fall within the limits X_1 and X_2. Failure of the putative percentage of terms to fall within these limits may be taken to indicate the presence of an "extraneous cause"—a cause not associated with the constant cause system identifying the random operation by which the past n terms of the sequence were generated.

The need for predictions in the form of probability tolerances as a basis for action is recognized much wider than it was a few years ago. Examples of action that are supposedly based on probability tolerances can be drawn from almost every field of scientific endeavor. For instance, no measurement, no survey, would be considered worth publishing unless there is evidence that repetitions would duplicate it within stated limits. Take government planning based on census returns; action can be based on the returns only if the census (whether by complete count or by sample) is carried out in such manner that repetitions of the survey, on or near the census date, would result in tables close enough to those that are actually published by the Census—*i.e.*, close enough for the purpose intended. The standardization of a drug is a problem in probability tolerances; dosage (future action) is based on the potency of the drug as determined by experiments which are carried out to make it possible to state in advance the limits within which 95 or 99 per cent. of the next 1,000 or 10,000 tests (results of doses) will fall. Recommendations regarding the planting of next year's crops should be based on predictions in the nature of probability tolerances for future yields, in terms of certain operations (variety, treatment, fertilizer). The prediction of ability for the selection of personnel calls for probability tolerances.

When probability tolerances for future terms of a random sequence are to be set on the basis of past terms alone, mathematical statistics is involved in the pure state—a problem of Type A as I have elsewhere classified the problems of statistical inference.[2] When the sequence is not random, the problem is not entirely mathematical, and may even be hardly at all mathematical; recourse must then be had to whatever knowledge there is concerning the underlying causes that make the terms of the sequence behave as they do. The problem is then one of Type B. The problem is partly or mostly Type B when there are not enough terms in the past, even though they are random, to allow probability tolerances to be set close enough for the purpose by mathematics alone. Often there is only one term, one result, one experiment; the rest is knowledge of the subject-matter.

THE SHEWHART CONTROL LIMITS ARE ACTION LIMITS; THEY STRIKE AN ECONOMIC BALANCE BETWEEN TWO KINDS OF ERRORS

Briefly, the Shewhart methods consist of studying a sufficiently large quantity of data, not as a single sample, but in a sequence of rational subseries that are obtained by breaking up the data in a manner that will indicate with a minimum amount of data the presence of extraneous causes, and the existence of stability when stability exists between the subseries. The most useful way of breaking up the data into subseries must be determined in each problem by knowledge of the subject-matter, but there are some guiding principles to keep in mind. It is important to collect the data and plot the points without delay, in order that action can be taken in time to ward off trends that are not wanted.

The presence of extraneous causes worth looking for is indicated by failure of a point on the control chart to fall within the "control limits." The problem of setting probability tolerances and control limits is a problem in action: What to do? to look or not to look for an extraneous cause? There must be an economic balance struck between the two kinds of errors—looking for trouble that does not exist, and failing to look for trouble that does exist.

There is not time here to tell you how to go about calculating and plotting a control chart and how to draw the control limits. I can use my time to talk about principles. The routine is simple enough, but a good many hours are required to become familiar with it. Fortunately, with the publication of the American War Standards,[3] and Colonel Simon's book,[4] it is possible now to point to a pretty full account of the steps in application.

[3] See references further on.

[4] Col. Leslie E. Simon, "An Engineers' Manual of Statistical Methods," John Wiley, 1941.

WHAT OPPORTUNITY IS THERE FOR MATHEMATICIANS?

In talking to a group of mathematicians I should like to seize this opportunity to point out where future emphasis in mathematical statistics must be directed if the present needs of the country are to be met. The accompanying diagram shows what part of the field has been tackled, and what remains. Most mathematical statistics has so far dealt with the central part of this diagram, where randomness exists—statistics *in vitro*. There are exceptions; Professor Wilks,[5] for instance, has made excursions of far-reaching importance into the right-hand part.

Randomness attained here ↓   Present ↓

Past		Future
Randomness being attained here, perhaps requiring months. The Shewhart methods assist, but so far there is but little mathematics, though there is some (theory of runs; the Wallis-Moore test). More mathematics needed.	Randomness exists here. Statistical distribution theory applies. Here is where most of the mathematical work has been done. More needed.	Predictions required in the form of probability tolerances, as a basis for action, and for identifying extraneous causes. The Shewhart methods work; also the mathematical methods of Wilks. More mathematics needed.

DIAGRAM OF THE STATISTICAL STUDY OF CAUSES

The practicing statistician welcomes all the distribution theory that has been developed to date. Far from there being too much mathematical statistics, even in the state of randomness there is not enough! With more time I could describe, for instance, some fascinating unsolved problems in distribution theory of great economic importance, and of extreme difficulty, that have arisen in the work of the Census.

Mathematicians have too often assumed that it is easy to produce random variates. It is easy to assume that, and the problems are interesting. Of course, not everything is or should be useful, and we should be thankful for that, but the usefulness of much mathematical work in statistics has often been greatly overestimated because randomness is presupposed, with no indication given to show how far the calculations are invalidated by lack of randomness and what to do in the absence thereof.

Urgent requirements of government and industry now challenge the best mathematical talents in the country. In spite of the volumes of texts and periodicals that have appeared in the statistical field, there are parts of the subject yet untouched—one could almost say yet unrecognized. The need is urgent: the opportunity is great. New tools, new channels of thought are being called for. For one thing, the new mathematical developments in the left-hand side of the diagram must take account of the *order of the observations,* as is already done in the theory of runs and in the Wallis-Moore test, for example.

The requirements of the present and immediate future, when recognized, may turn out to be the sun that draws forth the best efforts that mathematicians have to offer.

OTHER FUTURE REQUIREMENTS

In the future there will be increased attention to the operational point of view, by which I mean that statements made by statisticians will of necessity be confined more and more to statements that theoretically at least can be subjected to test, because their statements will be tested.

The Shewhart methods were devised for the two extremes, right and left, in the diagram. The Shewhart methods work! In the future, mathematicians will show why these methods work and will then set out to improve on them. Meanwhile, it is important that men in government and industry be taught how to use them (*cf.* the last section).

Theoretical statisticians have too often evaluated the methods of quality control in terms of their own theories, without stopping to acquire an appreciation of the problems that face the practicing statistician, who must stand by and see his predictions put to a test. For instance, I have seen it argued (innocently enough) that people in quality control ought to use the nomenclature and language of the theory of estimation.

There will be increased attention to the distinction between chance as defined *a priori* as a basis for theoretical calculations in probabilities, and the meaning of the word chance when it is used to designate relative frequencies predicted to fall within certain limits in a controlled experiment. Calculated chances are one thing, but actual observed frequencies, even under a constant cause system, may be another. In other words, it is necessary to recognize the distinction between what is in the bowl, as determined by some standard method of examining part or all of the chips, and what we get by performing some other operation on it. One can not set probability tolerances regarding the results of drawing from the bowl, purely from knowledge of what is in the bowl as determined by some standard method.

For example, the phenomenon of the undercount of children under 5 is well known among sociologists.

[5] S. S. Wilks, *Annals of Mathematical Statistics*, 1941: pp. 91–96. At the Vassar meeting Professor Wilks expounded and extended his published work.

Whether the census is taken by complete count or by sampling, the undercount of young children makes its appearance. How do we know? In the same manner in which it was discovered many years ago. In the census 10 years hence, with allowances for deaths, and for in and out migration, there will be more children of age 10–15 in any sizable area than there were children of age 0–5 ten years before. The phenomenon invariably repeats itself, in this country and elsewhere. People just will forget the baby when the Census man comes!

In this illustration, the count 10 years hence is the "standard measurement." The census method of counting the children of age 0–5 at the date of the first census is the other operation. If it were repeated at small intervals of time, it would perhaps show statistical control. Suppose it does: neither the average result nor the probability tolerances for this "other" operation can be determined *a priori* by the standard count. Likewise the converse: it is too much to hope that the standard count could be determined by the average result or the probability tolerances for the census operation of counting children of age 0–5.

SOME REMARKS ON THE PRACTICE OF STATISTICS

Faced with realities, it is sometimes difficult for the practicing statistician to frame his problems in the language of the theoretical statistics that has so far been developed. The problem of attaining control, for example, is not analysis of variance, is not a test of significance, is more than a problem in probabilities, and more than a problem in the statistical testing of hypotheses in the usual sense. It is a problem in *action*. The control limits must provide an answer to the question: What to do? Shall we or shall we not look for trouble? An economic balance must be struck between two kinds of errors (*supra*). The question is not whether an extraneous cause is operating to make a point plotted from inspection data fall where it does (inside or outside the control limits): the question really is *whether it pays to look* for an extraneous cause (trouble).

WHAT CAN BE DONE?

Now is the time to do something. Statisticians have a great deal to offer. Let them bring it forth. The following 7-point plan is offered as a starting point, to be replaced by better suggestions as experience develops.

i. *Intensive courses in quality control.* By intensive I mean intensive, covering 7, 8, or 10 days, including Sundays, 8 hours per day. Stanford University gave such a course in July, and another in September. Thirty-two picked delegates from industry attended in July, and 31 in September. They were not just college students or graduate students who might some time go into industry and have use for the course. As I say, *they were already in industry.* Most of them were well advanced in their organizations, with authority to try out statistical methods, and they went home with the ability to do so. The 80 hours' instruction was sufficient to give them an excellent start. All phases of the subject were taught, including diagnosis of data brought to the conference, and arrangements for continuation study afterward. The instructional staff at Stanford was supplemented by practical men from government and industry—a vital necessity in such undertakings.

Something similar is being done in Great Britain, and can be done by other institutions of learning in this country. A semester course in quality control is good so far as it goes, but it is not sufficient for the present needs. It is too slow, and the yield too low. The important thing is to give intensive courses to men in industry, now, patterned after the ones at Stanford.

The Ordnance Department some months ago commenced an educational program in quality control under the direction of Mr. G. D. Edwards, of the Bell Telephone Laboratories, who with Mr. H. F. Dodge, of the Bell Telephone Laboratories, and Mr. G. Rupert Gause, of Colonel Simon's staff at Aberdeen, are holding three-day schools for ordnance people in a number of key cities.

With the close relation between price control and quality control, and the necessity for quality determination and the standardization of textiles, foods, drugs and hospital supplies, the need for expert statistical work in this line is multiplied manyfold under present conditions.

The Farm Management Statistical Clinic held under the guidance of Professor H. C. M. Case at Urbana in September, 1941, was an example of a step in this direction.

ii. *Intensive courses in other fields.* Fields other than quality control could be covered by intensive courses for the benefit of men in government and industry. For example, something could be done in the way of teaching statistical methods in psychological tests for placement of personnel.

iii. *Modification of instruction in regular courses.* In regular courses of instruction, students have often been given an incomplete picture of the problems of mathematical statistics. The middle part of my diagram has been taught, and rightly so, but the student has not realized that he was being taught in only this limited portion. The result has been over-confidence in his ability, nonsense calculations and outright mistakes that might be serious if not caught, with consequent mistrust of statisticians. Of course, it is impossible to equip a student of statistics with all that he is going to need, and to give him full directions about just what to do in every situation. Every problem requires intricate tailoring of general principles. This will always be true, but the teaching of general principles has not been broad enough, and what is worse, the student has not been advised and warned of his deficiencies.

iv. *Short monographs on various statistical procedures.* All over the country I have found men in industry, in government and in research organizations eager to learn the tools that the statisticians have developed. From everywhere comes the question: "Where can I learn more about this? Is there anything printed in simple language that I can understand?"

Here again, something can be done, and is being done. I can refer you, for instance, to the American War Standards in the field of quality control. With remarkable foresight, perceiving the coming need for statistical methods in war industries and ordnance inspection, the War Department in December, 1940, asked the American Standards Association to develop concise treatments for the application of statistical methods, for use both in acceptance inspection, and during manufacture. A committee was appointed, with the result that two standards appeared in May, 1941, and a third in July, 1942, with titles as shown below:[6]

1. Guide for Quality Control, 1941
2. Control Chart Method of Analyzing Data, 1941

(Nos. 1 and 2 are bound together and sell at 75¢)

3. Control Chart Method of Controlling Quality During Production, 1942: 75¢.

These serve as guides in application and for introductions to texts and other books on the subject, and they contain references to texts and monographs on the subject for further study. Similar brochures on procedure could be produced for other statistical techniques.

v. *References to current articles and books in various fields.* The Institute of Mathematical Statistics and the American Statistical Association could perform a valuable service by running in each issue of their journals a list of references to current articles and books on applied statistics, broken down into various fields—sampling in social and economic surveys, quality control, psychological tests, presentation of data, and other subjects. Promptness is more necessary just now than completeness. Work should be commenced at once, with a three months' deadline for the first list, which should be an attempt to show important articles and books that will be helpful to statisticians in practice. The list would be kept current, and omissions filled in, by lists in subsequent issues.

Not all statisticians take the journals. The lists should be offprinted and advertised for sale separately: people would be glad to pay for them; the question is where can they buy them? Offprints of this nature, readily obtainable, would be excellent advertising for both the association and the institute, because they would show the public that these organizations are alive to present statistical needs and are perhaps worth belonging to.

Here is a splendid opportunity for some one not already overburdened by war work to get into it. The project could even be expanded, by the right person, to provide abstracts or films of foreign or rare articles worth looking at.

vi. *Exposition of methods in published articles.* In both journals, there could be more serious attempts to show the reader how to apply the methods that are presented. No matter how theoretical an article may be, it should nevertheless describe how its contents affect present practices. It should contain directions for application, plainly marked, "Step 1, Step 2, do this, do that, . . ." The hypotheses and conclusions and logic should be described in non-mathematical terms, as well as by strict mathematics. In other words, when a writer has something useful, he should say so, and make it possible for others to see why and how to use it. Incidentally, he should not expect the editor to do this for him.

vii. *Expert assistance.* This is the most vulnerable of my suggestions, but I think that some good might come of it, even if it leads to numerous failures. The fact is that there are not enough good men in applied statistics nor in mathematical statistics. There never have been. Right now, all over the country, there are young men grasping, groping, struggling with statistical problems on which action will be taken on a deadline. To whom can they look for assistance? If it could be made known, and if we were prepared, the association and the institute could assist by acting as a clearing house for statistical questions, by referring them to the right man, or by sending the names of experts who can and will go outside of Washington, New York, Chicago, and a few other places, and get their hands dirty. Expenses and a consulting fee would of course be paid; that presents no difficulty. What we need is some one in one or both of the organizations to take active charge of this project.

OBITUARY

EARLE RAYMOND HEDRICK

PROFESSOR EARLE RAYMOND HEDRICK, vice-president and provost of the University of California at Los Angeles, emeritus, died on February 3, 1943, at Providence, R. I.

He was born at Union City, Indiana, on September 27, 1876. He attended the University of Michigan (A.B., 1896), and was Parker Fellow from Harvard (A.M., 1898) at the University of Göttingen 99–01, receiving his Ph.D. in 1901. He attended the École Normale Supérieure, Paris in 1901. Dr. Hedrick was an instructor in mathematics at Yale University from 1901 to 1903, and professor at the University of Missouri from 1903 to 1924. He was professor of mathematics at the University of California at Los Angeles from 1924 to 1937, and served as provost and vice-president from 1937 to 1942. Since October, 1942, he was visiting professor at Brown University.

Throughout his life he was always greatly interested in the teaching of mathematics from secondary school to graduate school. In collaboration with the late C. A. Noble he translated the first third of Klein's "Elementarmathematik vom höheren Standpunkte

[6] Published by the American Standards Association, 29 W. 39th Street, New York. Reissued in Great Britain by the British Standards Institution, 28 Victoria Street, London, S. W. 1. The committee consisted of H. F. Dodge, Leslie E. Simon, W. Edwards Deming, Ralph Wareham, A. G. Ashcroft and John Gaillard.