From *In This Issue*

Dorian Shainin

While serving as Chief Inspector of the Hamilton Standard Division of the United Aircraft Corporation, Dorian Shainin became more and more concerned about the lack of sensitive and objective means for the solution of industrial problems. In his search for an adequate approach he found himself going back to the notes of a course in statistics he had taken at the Massachusetts Institute of Technology.

"Each year that my associates and I combined the work of others with our own developments, I became more confident that really difficult problems beyond the scope of my department - and even the company - would be considerably simplified by a statistical approach," Mr. Shainin explained to the Editors. Consequently he left United Aircraft some five years ago and joined the industrial consulting firm of Rath & Strong, Inc., in order to develop and implement his idea.

The article entitled *The Statistically Designed Experiment: A Tool for Process and Product Improvement* is the direct descendant of Mr. Shainin's original ideas, nurtured by his continuing experience with the concept over the course of the past few years. He has applied the statistical approach in numerous industries, including paper, printing, textile, rubber, silverware, clocks, and nuclear energy, and he has written extensively. In addition he has played a leading role in the American Society for Quality Control, and has received its Brumbaugh Award for outstanding contribution to the industrial use of quality control.

The Statistically Designed Experiment

Why run the risk of a bad guess about the causes of variations in quality when the factors in the problem can be persuaded to tell on themselves?

By Dorian Shainin

- In a metalworking concern, where a polishing operation follows plating, 90% of the units pass final inspection, but 10% are rejected. Why?
- A complex pneumatic device shows exceptional unit-to-unit variation in performance. Why?
- Units of a hydromechanical control mechanism vary widely in performance, even though most of its components are being kept to close tolerance. Why?

Making improvements in products or processes can be one of the most challenging - and also one of the most frustrating - tasks confronting management. To solve a single problem, numerous hypotheses may be advanced, tested, and found wanting. Experiments may drag on for years. After every approach has been exhausted, the company still seems to be up against a series of endless and ever-changing variables. If only it were possible to find the right combination, control the right factors, maintain the proper balance, so as to achieve continuous, economical, trouble-free operation!

Is there no better way than trial and error to solve such problems? Cannot the traditional approach be bettered? What is needed is a method that gives assurance that when a change is made in the product or the process, it will be the right change. The improvement created should be adequate and lasting. It should not cost too much or consume too much time.

An approach that fulfills all these requirements is now available in *statistically designed experiments*.[1] Already these have solved a variety of problems in many industries. The purpose of the present article is to indicate briefly what this technique is, what advantages it offers over conventional methods, and what kinds of problems it can solve.

Basic Features

The essential feature of the most up-to-date statistically designed experiment is the *simultaneous* consideration of a large number (sometimes all) of the possible causes for a product or process problem. It can *categorically* rule out most of the possible causes after a *limited* number of experiments. This means that the major source of trouble can be more and more closely pinned down until it is finally isolated.

The approach often makes use of but *never depends on* hunches and guesses in problem diagnosis. If the initial hunches happen to be right, the time for the experiment may be cut down; yet, if the hunches are wrong, as only too often happens the experimenters' efforts are not held up until new hypotheses can be formulated. Because statistical design can impartially evaluate most or all of the causes of a problem, it is a completely *objective* device.

Finally, acceptance or rejection of hypotheses and consideration of alternatives can be evaluated in terms of *known confidence levels*. The risk of wrong decision can be reduced, for all practical purposes, almost to zero.

Drawbacks & Advantages

Following from these basic features are a number of pros and cons worth noting. On the adverse side, it must be admitted that exploiting the new technique is not a do-it-yourself proposition - at least for companies without a well-qualified statistical engineering group.

[1] Sir Ronald a. Fisher laid the foundation for this approach in *Statistical Methods for Research workers* (London. Oliver and Boyd, Ltd., 1925) and in *The Design of Experiments* (London, Oliver and Boyd, Ltd., 1935).

Knowing how to apply these procedures calls for training; knowing which ones to apply in any particular situation calls for experience and judgment. Accordingly, large companies may wish to develop a group for this activity; smaller ones can get along with one or two people who could combine statistical work with other kinds of duties. Consultants could be hired to do the training.

Granting that the need for specialized help may be a drawback, what are the countervailing advantages? These seem substantial, and they should appeal to businessmen who put a high premium on time as well as on money:

(1) Usually the time involved in problem solving by statistical methods is short. It is not impossible for a single week's investigation to come up with an answer that has been eluding a company for years.

(2) Quick problem diagnosis, leading to quick cure, cuts the expenses of a high-cost operation, and the savings realized quickly pay back the initial costs.

(3) The process does not disrupt production; only minor interference is usually necessary. The experiments can be brief: the tests often involve only a relatively small number of units: and tinkering with operating methods on the line and with product specifications is kept to a minimum.

(4) Through training concurrent with problem solving, the company becomes progressively better able to carry on statistical activity at no extra cost with its own personnel.

Because this approach is fast and inexpensive, it is practical not only for well-heeled industrial giants but also for middle-size and small concerns. Thus, unlike many current developments in business techniques, the use of statistical procedures helps rather than hinders the competitive chances of small firms.

Underlying Logic

The theory and practice of statistical design rest on a series of simple logical propositions. Although these are not all self-evident, I think they will strike anyone who has ever worked on product or process problems as valid:

(1) Every effect has one or, more often, a number of possible causes.

(2) When there are many possible causes, the major portion of the effect usually comes from one or, more likely, just a few causes.

(3) These few major causes are not constant in their activity: they produce variation in the end product (the effect).

(4) Therefore, if variations in the end product are analyzed and related to their possible causes, one factor (or part of the total variation) may be expected to show up as being more important than the others, and the unknown cause may be associated with that particular factor.

On the basis of this reasoning, it is a fairly straightforward matter to design an experiment that will enable management to isolate and evaluate the reasons for undesired deviations from standard in a product or process. The action - the way the parts of the total variation show up - is made to "tell on itself."

A few case examples will serve to illustrate the possibilities.

A Quality Problem

Let us start with a problem of product quality - one that had bothered a metalworking firm for years, although it proved simple and quick to solve once an objective statistical design procedure was developed:

After a final polishing operation, the company experienced what was believed to be an excessive rate of rejects and consequent expenses for reworking. Polishing followed plating, and too often the hand-held polishing wheels exposed the base metal. Plant supervisors were certain that the difficulty stemmed from variation in skill among the several polishers and/or some inherent differences in the cloth from which the polishing wheels were made.

Believing that the supervisors' hypotheses sounded reasonable, the statistical engineer assigned to the problem decided to begin by determining, with a given statistical confidence level (a fixed per cent of certainty), who were the least and the most successful workers; then a study could be made of their different polishing habits, the results of which would be incorporated into a training program. In addition, the engineer planned to compare polishing wheels in order to see if variations in their material bore any relationship to work quality.

For statistical validity, it was necessary to run a short test during which a random mixture of different parts would be issued to each polisher, and each man would use one polishing wheel. A protective statistical "level of significance" was chosen so as to distinguish between chance variations in output and real differences arising from unequal skills. Interestingly enough, the results of this test showed *no significant difference* among operators. In other words, the major cause or causes being sought had been distributed about evenly (randomly) among all the polishers.

This unexpected revelation brought on a decision not to guess further about the major causes of trouble without some really objective evidence to indicate their nature.

Objective Inquiry

Freeing his mind of all preconceptions about the problem, and thinking back to the four logical steps outlined above, the statistical engineer now decided to look at the plating process (in contrast to the polishing operation, which had been pretty conclusively proved innocent); to consider the total variation in plating thickness in terms of three of its factors; and then to let the action of these factors tell which one of them was the most important and thus where the cause of most of the variation lay.

The engineer divided thickness variation into that occurring (a) from time to time, (b) from plating tank to plating tank, and (c) within a tank. Following this line of inquiry, a simple experiment was performed:

- Parts were identified for a short period according to the side of the tank and tank number in which they were plated, and according to the hour when the plating was done. A small variation in plating thickness showed up in the hour-to-hour figures and in the tank-to-tank figures, but none of the parts plated on the right-hand side of the tanks had the plating polished off, while many of those from the left-hand side had thin plating and had been rejected for exposure of base metal. Therefore, something that correlated with "within-tank" variation had to be controlled to move the variation in thickness toward zero.

- Discussion with the plating foreman brought forth no clues. Anode to cathode distances and electrical potentials had all been balanced when the tanks had been installed. It seemed desirable therefore, to observe the plating procedure. The only nonsymmetrical feature seen was a hand valve on a pipe on the right-hand side of each of the 14 tanks. This pipe carried steam along the length of the tank at the bottom on the right, across the front, and back on the left, rising up and out. The steam kept the plating solution warm, which was necessary for good results.

- A reason for polishing difficulty now became clear - a reason that fitted the observed facts. The steam must be hotter on entering the tank than on leaving it, so that the right side of each tank was warmer. Since warm water rises, a counterclockwise circulation of the plating solution must have been created. That meant that the plating particles coming from the anodes were in a rising current on the right side and in a falling current on the left. The articles handing on the left side of each tank, therefore, must be getting less thickness of plate.

- The steam valves were then closed. Parts from both sides of all tanks were polished. None were rejected.

Solution Found

The solution to the problem, therefore, was to relocate the heating coils in the tanks in order to avoid a circulation that would affect plating thickness.

Thus, a major quality problem *told on itself* in less than one week. The answer unfolded as soon as the statistical engineer insisted on an entirely objective approach, unbiased by what the management "knew" to be the crucial factors - operator skill and/or polishing wheel differences. As is often the case, the solution to the problem was almost ridiculously simple once the components of variation were carefully studied.

A Performance Problem

Solving the polishing problem just outlined was relatively simple; a slightly more complex statistical design may have to be worked out for other problems. A representative case in point involves a manufacturer of a complicated pneumatic unit:

- An inexplicable unit-to-unit variation occurred in the production of this complex item. For proper operation, some units required considerably higher supply-line pressure than others. This situation was unacceptable to the customer and had been under intensive investigation for months, while company engineers theorized as to first one, then another, possible cause. So far, no changes in dimensions, tolerances, assembly, or test procedures had resulted in the improvement desired.

- Feeling that the clue to performance variation lay in some unknown physical differences among the units, but having no idea what these differences might be, the statistical engineer selected two units from a day's production which evidenced the greatest variation in the pressure required for satisfactory operation. One unit was tagged as the "Low" unit, the other as the "High" unit.

- After a discussion with the engineers on the job, two subassemblies common to both units were selected as possibly accounting for most of the variation in final unit operation. These subassemblies were removed from the units and identified as "A" and "B," while the rest of the unit was identified as "R." The six components ("A" from "High" and "Low," "B" from "High" and "Low," and "R" from "High" and "Low") were then reassembled and tested in the following combinations:

A-High with B-Low and R-High
A-Low with B-Low and r-Low
A-High with B-High and R-Low
A-Low with B-High and R-High

- Tests were run on each combination *twice*, giving eight test runs in total. These eight test runs were performed in *random sequence*, to minimize the possibility that chance environmental factors affecting test conditions could "throw off" interpretation of results.[2]

- Next, the readings for each test run (in pounds per square inch) were entered on a diagram, known as a "Latin square," opposite the test number:

	A-High	A-Low
B-Low	(1) 20.40 R-High (2) 20.50	(5) 20.60 R-Low (6) 20.40
B-High	(3) 22.90 R-Low (4) 23.10	(7) 23.00 R-Low (8) 22.90

- By averaging the four readings in each vertical *column*, the effects of A-High and A-Low could be measured.
By averaging the four readings in each horizontal *row*, the effects of B-Low and B-High could be measured.
By averaging the four readings *diagonally*, two ways, the effects of R-High and r-Low could be measured. That is, the average of the upper left and lower right boxes shows the effect of R-High; the average of the lower left and upper right boxes shows the effect of R-Low.

- Significantly, because of this balanced Latin square design, each pair of averages reflects the difference caused by a change in one particular variable. The effects of the other two variables, while included, are exactly balanced and are therefore neutralized.

Clues to Solution

For evaluating the results of statistical experiments such as this one, special methods have been devised to indicate within a known confidence level (i.e., acceptable margin of error), whether the differences obtained among average figures are large enough to be significant or not. In this case, on examination of the data, it was apparent even without a statistical test of significance that subassembly B was responsible for the relatively large variation among units. The average results of B-low and B-High showed too great a difference to be attributed to chance, or to the influence of other variables which caused differences among the results within each box.

The results of this experiment suggested a careful inspection of the B subassemblies from both the Low and the High units. A dimensional difference was found to exist in the length between a fulcrum and an actuating point on an arm - a difference which could easily be controlled in future production.

Thus, once more, when variation was broken down into its components, immediate clues to the nature of the problem came to light. The company's previous experience had led it to believe that the solution lay in redimensioning subassembly A. Fortunately, this expensive and unnecessary course of action was avoided.

Other Possibilities

In this instance, one of the two subassemblies first selected for investigation turned out to be the source of the company's problem. What if this had not been the case? Under these circumstances, readings from the diagram would have readily disclosed what the next step ought to be:

(1) If the large difference in averages had remained in R, the test would have been repeated using C and D subassemblies rather than A and B. Eventually, this process of elimination would isolate the source of variation involved.

(2) If the large difference were found to be "within box" variations, this would indicate that test equipment or other conditions of environment were at fault. Such

[2] For a helpful discussion of randomization, see W. Allen Wallis and Harry V. Roberts, Statistics: *A New Approach* (Gencoe, Illinois, The free Press, 1956), p. 119 and pp. 479-481.

"experimental error" would indicate the need for a more carefully controlled test procedure.

(3) If only one of the four boxes had results out of line with any other, and the experimental error was small, this would be evidence of an *interaction* among two or more components. Or if one R-Low box had a pair of noticeably high readings and the other R-Low box had a pair of the lowest readings, that also would point to interaction. To track down a problem caused by interaction, the *"factorial"* design of experiment should be used instead of the Latin square.[3] The two methods are similar in principle, but the factorial design requires that every *possible* combination be run at least twice.

A Standardizing Problem

A third statistically designed experiment concerns a company's efforts to reduce variation in the output of a hydromechanical fuel control. In order to use this unit in a new application, a 50% reduction in output variation was required, which seemed to present insurmountable problems.

Most of the critical components and dimensions in the assembly, of which there were many, were already being held to closer tolerances than desired for economical production. Furthermore, trial-and-error methods had given no clue as to what changes might lead to the desired reduction of output variation. It appeared that a major redesign, or a new design altogether, was needed.

With no indication at all of what might be the cause of the trouble, the statistical engineer set about to design an experiment that would unearth it. There were literally dozens of critical "characteristics" (such as a hole diameter, a spring load, a concentricity requirement, a critical length, a radius, a spring rate) in the total unit. Any one of these might be at fault - or the fault might lie in an interaction among them. Furthermore, the "value" of each such characteristic could vary, depending on the tolerance to which it had been held in production.

In order to run a full-scale factorial experiment, testing and retesting several values of all these critical characteristics in every possible combination, the engineer would have had to make literally thousands of tests. He chose rather to take a random sample of these combinations - to conduct a more limited experiment in the hope of solving the problem faster and with less expense:

- He chose to run only 30 tests, *randomly* assigning to each a value for every critical characteristic. Not all combinations would be run; in fact, only a small percentage of the total possible would be tested. But there would be enough to constitute a good sample.

- For each critical characteristic he chose three approximate values, "High," "Medium," and "Low." When the characteristic was physically adjustable (e.g., a spring load), only one component part was required to make the tests. When such adjustments could not be made, available fuel controls often provided examples close to all three values. In certain instances, parts had to be manufactured to specific dimensions.

- The particular level of each characteristic to be incorporated into the assembly for each test run was selected by using a table of random numbers. The layout is illustrated in EXHIBIT I.

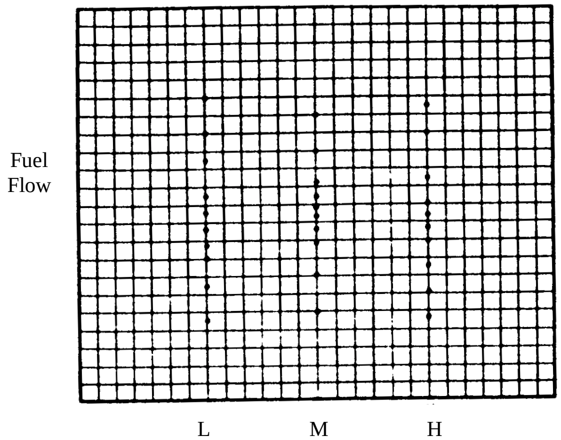
EXHIBIT I - Random Selection of Characteristics for Test Run

Test run sequence	Critical characteristics I	II	III	IV	V	etc.	Fuel flow in lbs/hr
1	M	L	H	H	L	. . . .	______
2	L	L	M	L	H	. . . .	______
3	L	H	L	M	M	. . . .	______
4	M	M	L	H	L	. . . .	______
5	L	H	L	M	M.	. . . . . .	______
.	.	.	.	.	.	. . . .	. . . . . .
.	.	.	.	.	.	. . . .	. . . . . .
.	.	.	.	.	.	. . . .	. . . . . .
30	L	M	M	H	L	. . . .	______

- Test run #1, then, would require that a fuel control unit be assembled that would include Characteristic I at Medium value, II at Low, III at High, IV at High, V at Low, and so on. A fuel flow was recorded for this control, and then this unit was disassembled and a second one was assembled to incorporate the requirements of test run #2. This procedure was repeated 30 times, and 30 fuel flow readings were recorded.

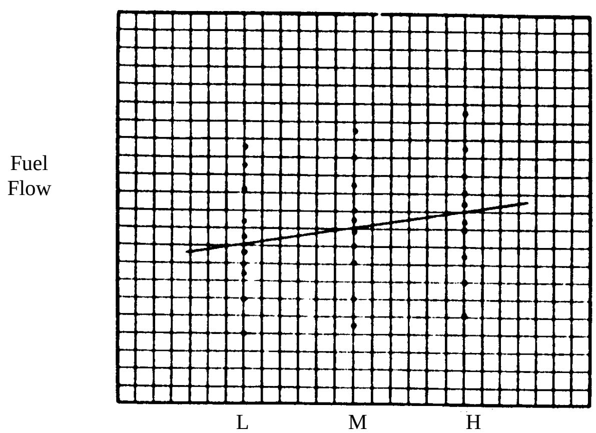
- To eliminate time-consuming mathematical calculations, the analysis of the results was performed graphically on ordinary squared paper. A group of 30-point scatter diagrams - one diagram for each critical characteristic - was drawn up, with the value of the characteristic plotted on the horizontal axis, against fuel flow on the vertical axis. For example:

[3] See W. Allen Wallis and Harry V. Roberts, op. Cit., pp. 480-481.



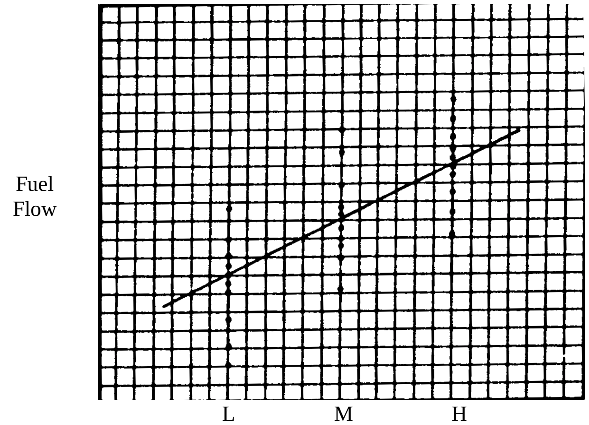
Characteristics II

In this instance, obviously, there was no correlation between the values of the characteristic and fuel flow. The tolerance specification for this characteristic was relatively unimportant, therefore. Now here is the diagram of another characteristic, which works out differently:



Characteristics IV

In this instance there was some correlation, but a wide scatter of points occurred about the trend line. The specification affected the results, but fuel flow was influenced much more by values of other characteristics which were entering the plot in a random manner and were the cause of the large scatter. So let us turn to another characteristic:

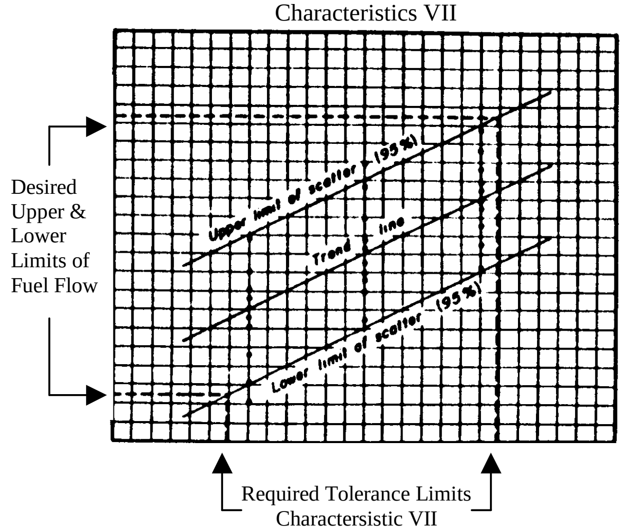


Characteristics VII

Here there is a good correlation, with a narrow scatter about the trend line. The specifications for this characteristic appeared to have an important effect on the results. In order to validate the finding, this correlation (as well as any other that appeared important) was checked by a few additional test runs in a bona fide factorial design, in order to establish valid levels of statistical significance.

- It was then a straightforward job to determine statistically the position of two lines, parallel to a trend line, which would include 95% of the points to be expected from a far greater number of runs. Next, from the desired upper and lower limits of fuel flow on the graph, a pair of horizontal lines were drawn (see EXHIBIT II).

EXHIBIT II. Computation of Required Tolerance Limits for Characteristics VII



The line from the maximum desired fuel flow stopped at the upper boundary of the scatter, that from the minimum at the lower boundary.

Perpendiculars were then dropped from these two intersection points. They showed, on the horizontal scale, the *real tolerance limits* required for Characteristic VII to guarantee (with 95% certainty) compliance with the desired fuel flow limits.

Solution Made Easy

This study provided the company with a completely objective evaluation of a highly complex design. Through careful study of the results of only 30 test runs, utilizing components from several fuel controls, it showed where tolerances were too tight, too loose, or just right - but sometimes out of position. Certain important and controlling tolerances had to be held closer in order to avoid the necessity of selective assembly. In some cases, specifications had to be changed to flatten the slope f the trend line so as to provide a practical manufacturing tolerance for those characteristics.

As a result of this experiment, the company was able to make the changes needed to bring the fuel flow within the desired limits *without major redesign*. In fact, only a few tolerances had to be tightened, and these were amply compensated by the discovery that several others were being held unnecessarily close.

Future Prospects

Statistically designed experiments are increasingly being used to solve production and process problems. But these limited uses by no means exhaust the potentialities of the new technique - potentialities that are still in the early stages of development and exploration and only await the passage of time.

A very large number and a wide range of problems can be attacked by means of techniques based upon the four logical steps enumerated earlier, i.e., letting the action *tell on itself*. For example, I know of two cases where the unknown but controlling factors in market activities were revealed; of another one in the paper industry where the predominant cause of variation stemmed from the natural, uncontrollable characteristics of trees (but could be compensated for); and of an extremely interesting application of the statistical design approach in the area of medical research.

Finally, it seems particularly significant to note that the usefulness of statistical designs is not limited to *problem* situations. These, after all, are relatively unusual in many industries. Far more common are situations where, for example, products and processes are adequate - but not so efficient or so economical as they could be. Now, through the use of statistical designs, companies at last have a relatively quick and inexpensive way to find out objectively where improvements can be made and what the specific improvements ought to be.

So, for the future, the possibilities for savings in cost and increased efficiency appear to be incalculable.