AQC93, 47th Annual Quality Congress, Boston, Massachusetts, May 24-26, 1993

STATISTICAL PROBLEM SOLVING STRATEGIES

H. J. Bajaria, Ph.D., P.E.
Multiface, Inc.
Garden City, Michigan

ABSTRACT

Statistical Problem Solving (SPS) strategies play a key role in connecting problem-solving approaches based on the principles of physical sciences and the mathematical efficiency offered by statistical science. Without the application of such strategies, practitioners of each scientific discipline seem to compete rather than complement one another. It is the purpose of this paper to discuss SPS strategies of attack combining both sciences. The real-life illustrations show their applicability and effectiveness in a variety of problem-solving scenarios.

SPS strategies of attack are discussed in three categories of problems: (1) natural variation problems, (2) instability problems, and (3) a combination of both types of problems.

This paper is intended to serve as a foundation to establish a need for strategies and to illustrate the variety of strategies available in solving problems effectively.

SPS STRATEGIES OF ATTACK

We begin by discussing the strategies of attack as shown in Table 1.

Table 1: SPS Strategies of Attack

Problem category 1	Natural variation	in series processes
		in parallel processes
		in processes with hierarchical process variables
		in processes with known awkward solutions
Problem category 2	Instability	Processes with unknown instability
		Processes with known instability
Problem category 3	Natural variation and Instability	Multivariate analysis for instability
		Multiple regression analysis for natural variation

PROBLEM CATEGORY 1 - NATURAL VARIATION

Natural variation in series processes

The best place to start problem solving for a series process is at the location where the problem is obvious and its measurement is not argumentative. Once such a spot is selected, then the first stage of investigation is defined as moving backward one step in the process. Since we are always seeking a robust solution (insensitive to variation in incoming material), the incoming material is always included as one of the investigative variables in search for the ultimate solution. Figure 1 depicts an approach for a series process consisting of three stages. We assume here that the output of process A and the output of process B cannot be measured in terms of the problem seen at the output of C. The implication of this assumption is that we must measure the success of any investigative attempt in terms of output at C. We begin with problem-solving attack at stage 1 which includes variables 7, 8, and output of B. If we fail to determine the solution at stage 1, we move up to stage 2 with a new set of variables. At this stage, variables are 4, 5, 6, 7, 8, and output of A. Upon failing to discover the solution at stage 2, we move up to stage 3. It is possible to gain incremental knowledge at each investigative stage and solve the problem partially as we move up the stages 1 through 3. Depending on the type of problem, the strategy discussed may slightly vary but the basic attack remains the same.

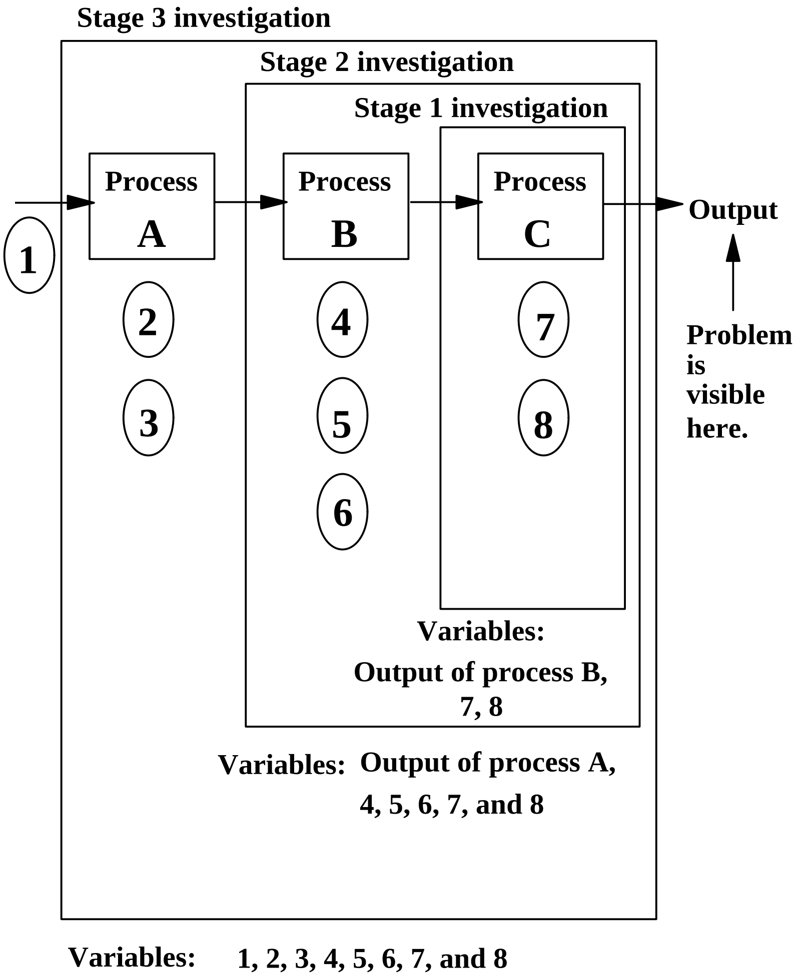


Figure 1 - Investigation stages for a series process

Figure 2 exemplifies how to approach a problem of crushed crackers and loose packages. The stage 1 investigation would include cracker height (incoming material variation) and spring force (packaging variable) as investigative variables. If the problem is not eliminated completely at this stage and if cracker height turns out to be a principal contributor, then stage 2 will move backward to the cooking oven. The investigative variables at this stage are dough properties, temperature gradient, belt speed, and cracker layout method. The effects of these variables on the cracker height will be investigated.

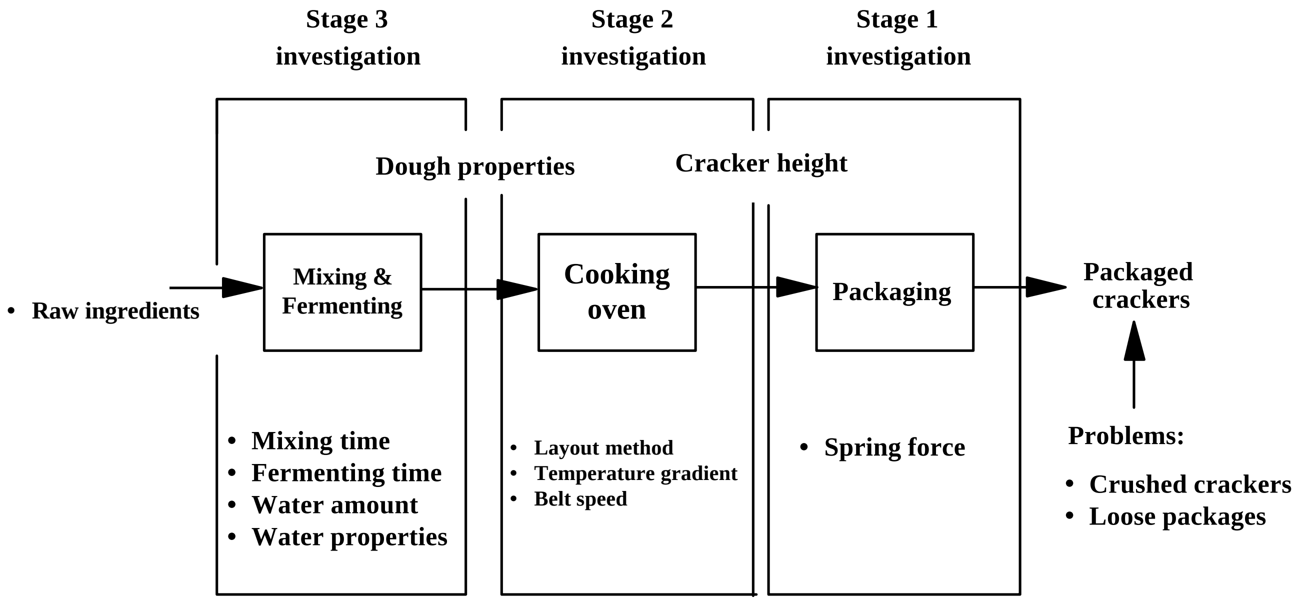


Figure 2 - An example of series process investigation

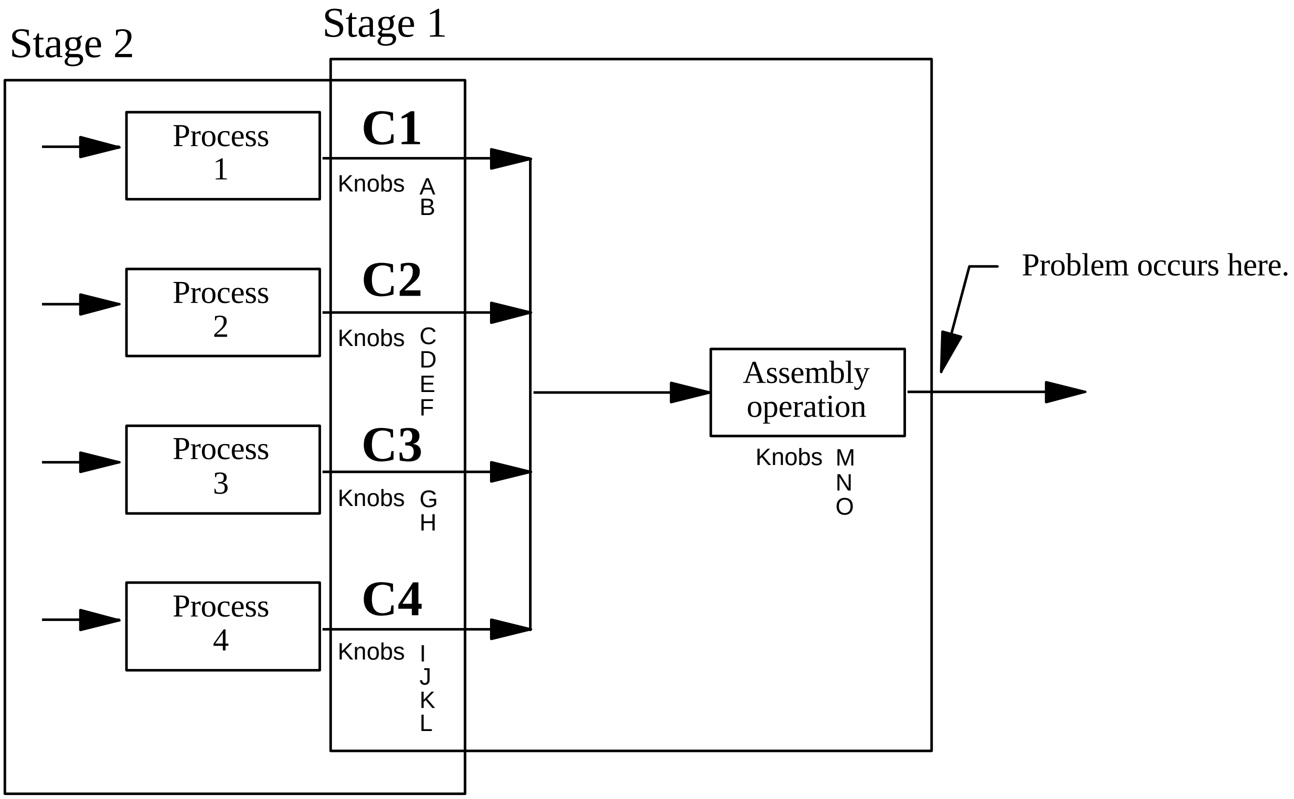
Once again, if the dough properties turn out to be the major contributors to the cracker height variation, then the investigation moves backward to stage 3. The investigative variables at stage 3 are mixing time, fermenting time, water amount, water properties, proportion of raw ingredients, and properties of raw ingredients. This problem is slightly different than a generic series process described earlier where we could not relate intermediate outputs to the problem seen at the end. In this example it is possible to measure the intermediate outputs which are dough properties and cracker height. Therefore, we can talk about solutions in terms of improving dough properties and cracker height.

Natural variation in parallel processes

The approach for parallel processes is similar to series processes with the following exception. The incoming material variable consists of multiple sources as opposed to a single source. An assembly process is an example of a parallel process. Figure 3 depicts an approach for problems with parallel processes. The complex assembly problems are resolved by this strategy.

Natural variation in processes with hierarchical process variables

In many investigative situations, the list of variables may be overwhelmingly long. However, if we begin to examine the possibility of arranging them in hierarchical order, the list may look considerably smaller and manageable. Figure 4 shows a generic hierarchical arrangement. The first level of hierarchy consisting of knobs X, Y, and Z is conceptual. For example, hydraulic, magnetic, electrical, mechanical, optical, and thermal are the concepts on which a process is designed. The second level of hierarchy is in the form of tunable knobs (process variables) indicating how the process can be adjusted to suit the circumstances. For example, temperature and cycle time both can be adjusted to create a desired thermal effect.



C1 = Component 1

C2 = Component 2

C3 = Component 3

C4 = Component 4

Stage 1 list would contain 15 knobs: the three knobs at the assembly operation itself plus the knobs associated with each component going into the assembly.

Stage 2 list would depend on the outcome of stage 1 investigation.

Knob = Process variable

Figure 3 - Investigation stages for a parallel process

The strategy here is to attack conceptual variables (top hierarchy) at the first stage of investigation. For example, we can investigate the thermal variable before we explore temperature and cycle time in greater details. To do this we can choose short and long thermal exposure as two levels of thermal variables. To execute short and long thermal exposure with the tunable knobs, we can choose level 1 as a combination of short time and low temperature and level 2 as a combination of long time and high temperature. Only upon proving that the thermal variable is a major contributor, should we be interested in the complete understanding of time and temperature. This approach is exemplified in reference (1), page 239.

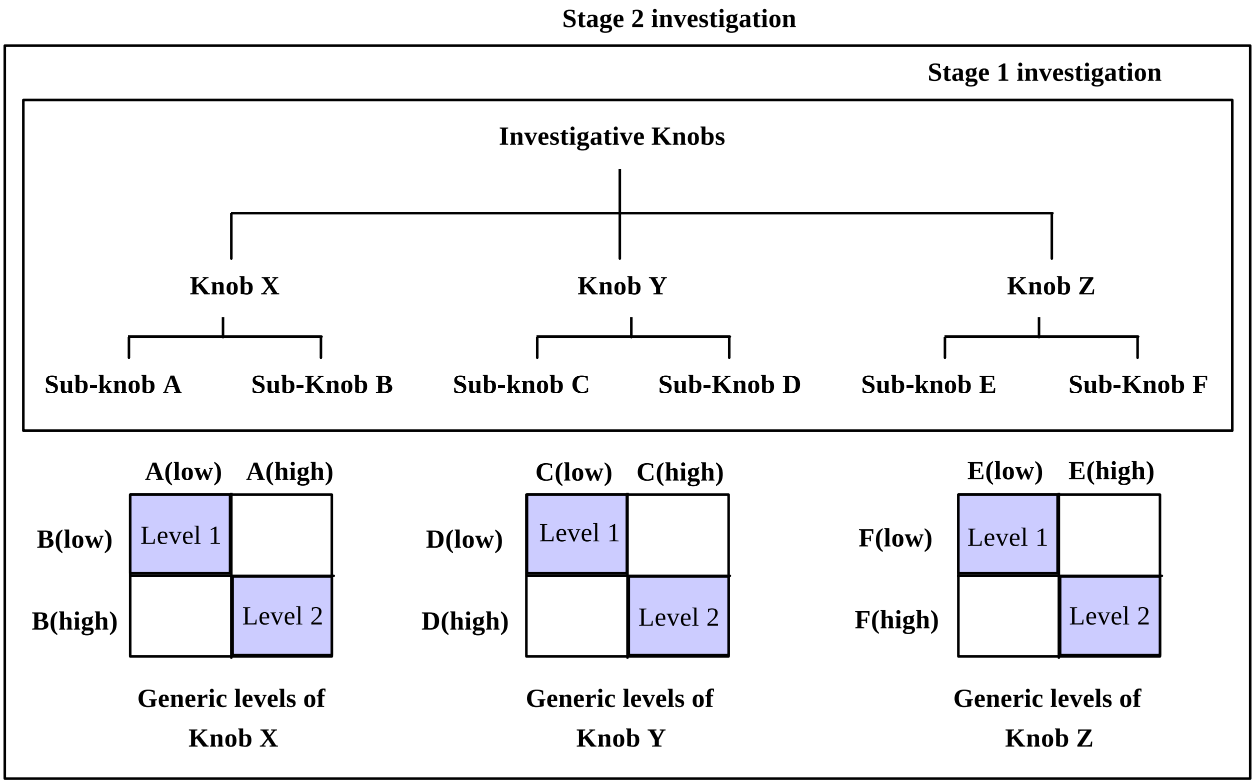


Figure 4 - A generic hierarchical knobs arrangement

Natural variation in processes with known awkward solutions

There is no need to go through an elaborate discussion of investigative variables when daily manipulation of variables eventually results in a smooth running operation. What is needed in such a case is formal experimentation to solidify and mechanize the knowledge elements that are already present. The most likely reason for this situation is an operator dislike for incoming material variation requiring excessive and ungraceful knob manipulations. In the conventional thinking process, this is called a *material problem*. What we call *material problem*, we deal with anyway. There are two fundamental reasons for defining some problems as *material problems*. These reasons are: (1) the process is designed with an assumption that the material properties are constant but in actuality they are not, and (2) the assumption that the process can compensate for incoming material variation, but such compensation is only manual and awkward. Nobody likes the idea of manual compensation for varying incoming material properties, and therefore these situations are labeled as *material problems*.

Here is an example of a problem with a known but ungraceful solution. A paper machine is pasting two papers together. The problem is that papers sometimes don't stick together. Some of the process variables that affect the sticking integrity are amount of glue, machine speed, paper temperature, glue viscosity, paper moisture level, glue temperature, and paper tension. The paper moisture level is the incoming variable and its moisture content varies between plus or minus 4%. In order to achieve consistent quality, the operators have to adjust the paper temperature and machine speed with every new roll of paper. The operators dislike the idea that every new paper roll requires different settings. Their struggle to compensate for moisture is labeled as a *paper problem* over the years. Thus, knowledge to achieve consistent quality is present but considered awkward requiring frequent adjustments. Formal experimentation reveals that the paper temperature can compensate moisture only up to a certain limit due to a limited boiler capacity. Beyond that, machine speed must be slowed to increase the heat-time exposure to derive the equivalent thermal effect. Once the formal relationship has been established among temperature, speed, and moisture level, a mechanization is installed as an inherent part of the process design. In the improved process design moisture is measured automatically, and compensation with heat and/or speed controls is also performed automatically. Thus, the problem is considered solved for all practical purposes. Only two significant things occurred in this example. First, the sketchy knowledge that existed over the years was formalized and second, cumbersome attempts were mechanized. The operators are now free to think about a higher order of problem solving.

PROBLEM CATEGORY 2 - INSTABILITY

Processes with unknown instability

If an instability is present but the root cause is not obvious, then an investigation must follow. The following statistical procedure is most effective in arriving at the root cause relationship. This is called divide and conquer strategy. The idea is to list all suspect variables. If these variables do not act simultaneously in creating a product, then arrange them in chronological order. For example, in an injection mold machine, if material is injected first, then heated up, and then held at temperature for some time, then a chronological listing would be material property, injection pressure, temperature, and cycle time. If the variables act simultaneously then the sequential order does not matter.

The next step is to divide the list of variables in half. Each half set is considered to be just one variable. For example, material property and injection pressure may constitute one half and temperature and cycle time may constitute the other half. In the first stage of investigation, two levels are selected from the first half set and two levels from the second half set. The selection of which two levels to choose is based on the engineering judgment, because they must represent realistic operational limits. There are six possible choices. Consider variables A and B in the first-half set. The six possibilities of two levels are: (A_1B_1, A_2B_1), (A_1B_2, A_2B_2), (A_1B_1, A_1B_2), (A_2B_1, A_2B_2), (A_1B_1, A_2B_2), and (A_2B_1, A_1B_2). The recommended choice is (A_1B_1, A_2B_2) due to A and B extremities giving the widest coverage to the investigative space. The purpose of this investigation is to narrow down the root cause region either to first-half set or to second-half set of variables. Let us say for discussion that among A, B, C, and D, set 1 includes variables A and B and set 2 includes variables C and D. The stage 1 investigation suggests that set 1 is a culprit. Thus the stage 2 investigation would be carried out with only two variables A and B. The stage two investigation would reveal either A or B as a culprit. Figure 5 shows the investigation sequence. In following this strategy, stage 1 required 4 attempts and stage 2 required an additional 4 attempts - a total of 8 attempts. Had this investigation been conducted by conventional DOE thinking, it would have required 16 or 2^4 attempts. Thus, this strategy saved 50% of the time.

This strategy would succeed if no interactions are present and only a single variable is the root cause of process disturbance. If at any stage we discover that the sets are interacting, then the strategy must be switched to a more suitable one.

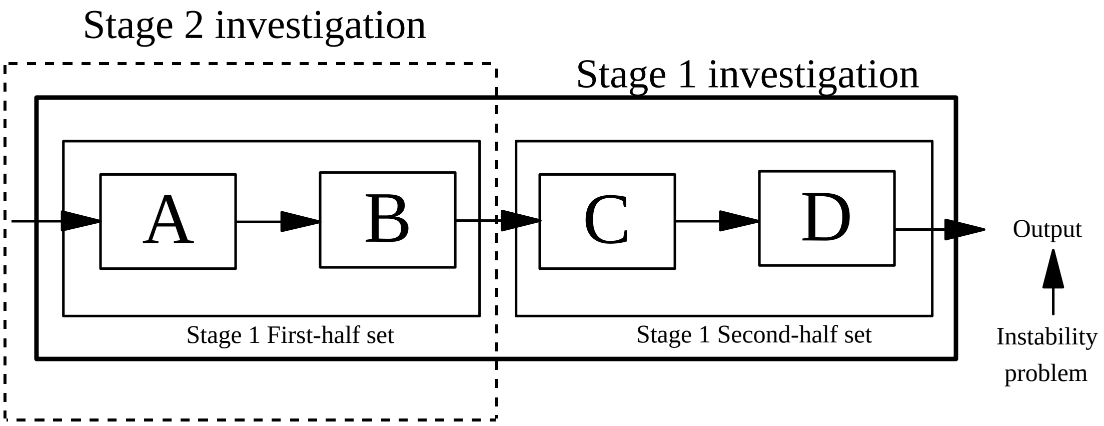


Figure 5 - Divide and conquer strategy

Processes with known instabilities

This approach works to investigate disturbances in a well-documented mature process. The most likely reason for instability to exist is the violation of process parameters. The best way to determine the root cause is to perform the audit of process procedures and parameters to reveal the most likely violation. The process control chart provides the clue for the inquiry of such violations.

Here is an example. This example relates to the same paper machine discussed earlier. It was noticed that the papers did not stick near the edges but they were sticking acceptably in the middle width. This was entirely a different problem than the moisture compensation problem considered earlier. The output control chart revealed that this edge phenomena got worse with respect to time. The audit of the process procedures showed that seeping glue near the paper edges was built-up on the rollers. The papers were riding on these rollers. The periodic glue cleaning was standard operating procedures but it was being violated. As the glue accumulated on the rollers, the paper shape became concave and the paper edges did not receive the same heat and pressure as the middle width. It was obvious that we must fix this procedure violation in order to solve this problem. Instead of manual cleaning, an automated cleaning procedure was installed - a change from operator control to a system solution.

PROBLEM CATEGORY 3 - NATURAL VARIATION AND INSTABILITY

Many times the output exhibits instability as well as incapability problems. Figure 6 depicts this scenario. Instabilities are due to physical disturbances in the process. Incapability is due to the interaction among the process variables. Multivariate analysis is used to understand those variables responsible for instability. Multiple regression is used to understand those variables responsible for incapability. The first step is to examine the output on an appropriate control chart such as shown in Figure 6. This chart exhibits signs of instability as well as natural variation. This form of instability can be appropriately labeled as *shift*, and to understand the root cause, multivariate analysis is performed. To understand the natural variation, we will have to perform multiple regression. Let us also say that for either problem to exist, three suspect variables, A, B, and C, have been recorded simultaneously with the output. We will now explore the details of multivariate analysis and multiple regression which can be used to understand variations seen in Figure 6.

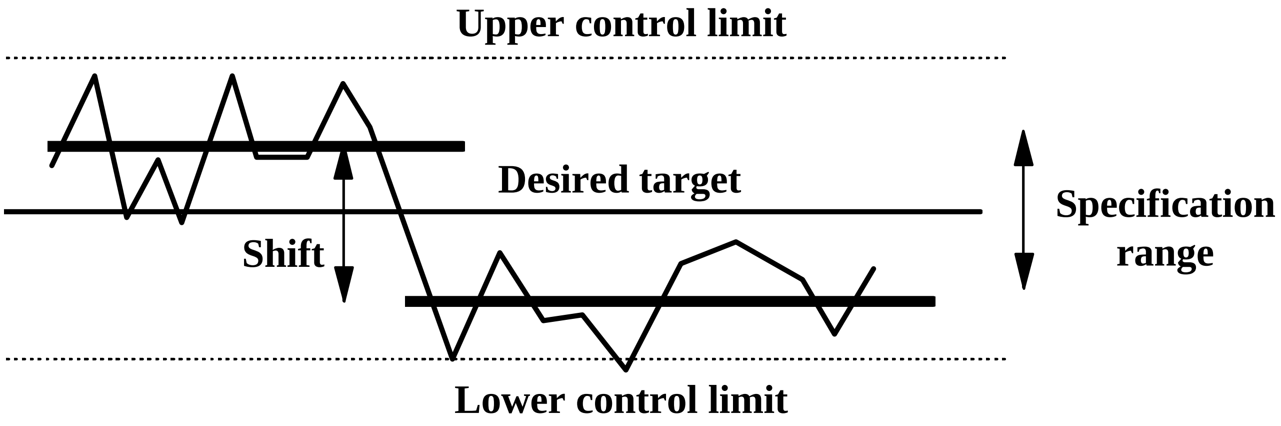


Figure 6 - Output control chart showing shift instability and natural variation

Multivariate analysis investigation for instability

A total of eight possibilities exist to analytically examine the reason for the shift in output. These 8 possibilities are: A, B, C, AB, AC, BC, ABC, and none. The first seven possibilities need to be plotted on a run chart to see if it matches the pattern seen on the output control chart. If a visual comparison shows that there is a match, then we have found the reason for instability. If we cannot find a match between the pattern seen on the process variable run chart and the output control chart, then we conclude that the instability is unexplainable by the seven possibilities, and we must continue to explore other reasons. Once we determine the reason for instability, we must implement a solution to control or eliminate that reason. This will eliminate the output instability.

To draw run charts for A, B, and C is straightforward. However, run charts for AB, AC, BC, and ABC must be drawn as multivariate run charts (Hotelling T^2 charts). Hotelling T^2 charts combine two or more variables into a single statistic. This single statistic T^2 then allows us to examine the patterns of two or more variables on a chart and match them with the output pattern.

Multiple regression analysis for natural variation

Once the reason for the instability is understood and removed, multiple regression analysis is used to analyze natural variation in the output. First, we can make a correction for the shift in the subset of the output data to make it look like a natural variation problem as shown in Figure 7. Then, we can use multiple regression analysis to understand the natural variation. Multiple regression is a process for fitting variables A, B, C, and their interactions into an equation to explain the output variation. To cover the details of multiple regression is beyond the scope of this paper. The reader should refer to an excellent reference for the detailed treatment on this subject[2].

An important point to remember is that multiple regression analysis should not be used blindly without examining the form of output data. It is easy to misapply user-friendly computer programs without thorough understanding of the multiple regression analysis sequence.

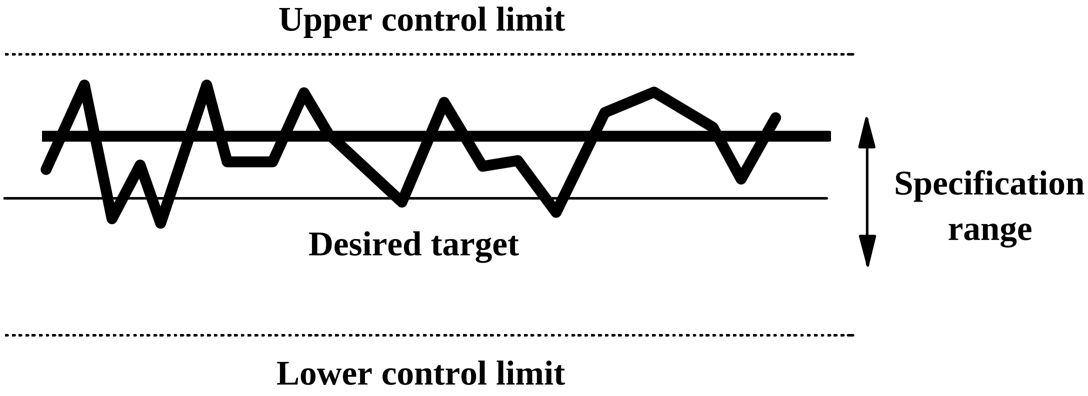


Figure 7 - Output control chart showing natural variation

SUMMARY

Problem-solving strategies are essential if we are to increase the knowledge of physical sciences by using the power of statistical sciences. These strategies effectively fill the fundamental gap in problem-solving processes. To an engineer, the best problem-solving scenario is to take actions dictated by the physical sciences with an intuitive sequence. To a statistician, on the other hand, the best scenario is to place investigative variables in the most mathematically efficient environment. This is a clear gap in the investigation mannerism. Additionally, statistical science cannot claim the knowledge of why a certain variable is being investigated, where the variable is physically located in the process flow, or whether two variables can be combined to look like one. To reduce this mannerism gap in somewhat of an acceptable manner, we must negotiate a compromising position.

The problem-solving strategies presented in this paper play such a negotiating role. These strategies are just as necessary as the principles of physical sciences and statistical sciences to fill the fundamental gap. Without these strategies, many solvable problems have remained untouched. And, those problems that do get solved without these strategies take longer time than otherwise. I hope this paper contributes an essential but often neglected question at every problem solving meeting: **"What is our strategy?"**

REFERENCES

(1) Bajaria, H., Copp, R., *Statistical Problem Solving: A Team Process to Identifying and Resolving Problems*, Multiface Publishing Company, Michigan, 1991.

(2) Neter, J., Wasserman, W., Kunter, M., *Applied Linear Statistical Models,* 3rd edition, Irwin, 1990.

LCS:521:10:000