4 Quality Engineering: The Taguchi Method

Taguchi's Quality Engineering Handbook. Genichi Taguchi, Subir Chowdhury and Yuin Wu

4.1. Introduction

The term *robust design* is in wide spread use in Europe and the United States. It refers to the design of a product that causes no trouble under any conditions and answers the question: What is a good-quality product? As a generic term, *quality* or *robust design* has no meaning; it is merely an objective. A product that functions under any conditions is obviously good. Again, saying this is meaningless. All engineers attempt to design what will work under various conditions. The key issue is not design itself but how to evaluate functions under known and unknown conditions.

At the Research Institute of Electrical Communication in the 1950s, a telephone exchange and telephone were designed so as to have a 40- and a 15-year design life, respectively. These were demanded by the Bell System, so they developed a successful crossbar telephone exchanger in the 1950s; however, it was replaced 20 years later by an electronic exchanger. From this one could infer that a 40-year design life was not reasonable to expect.

We believe that design life is an issue not of engineering but of product planning. Thinking of it differently from the crossbar telephone exchange example, rather than simply prolonging design life for most products, we should preserve limited resources through our work. No one opposes the general idea that we should design a product that functions properly under various conditions during its design life. What is important to discover is how to assess proper functions.

The conventional method has been to examine whether a product functions correctly under several predetermined test conditions. Around 1985, we visited the Circuit Laboratory, one of the Bell Labs. They developed a new circuit by using the following procedure; first, they developed a circuit of an objective function under a standard condition; once it could satisfy the objective function, it was assessed under 16 different types of conditions, which included different environments for use and after-loading conditions.

If the product did not work properly under some of the different conditions, design constants (parameters) were changed so that it would function well. This is considered parameter design in the old sense. In quality engineering (QE), to achieve an objective function by altering design constants is called *tuning*. Under the Taguchi method, functional improvement using tuning should be made under standard conditions only after conducting stability design because tuning is improvement based on response analysis. That is, we should not take measures for noises by taking advantage of cause-and-effect relationships.

The reason for this is that even if the product functions well under the 16 conditions noted above, we cannot predict whether it works under other conditions. This procedure does not guarantee the product's proper functioning within its life span under various unknown conditions. In other words, QE is focused not on response but on interaction between designs and signals or noises. Designing parameters to attain an objective function is equivalent to studying first-order moments. Interaction is related to second-order moments. First-order moments involve a scalar or vector, whereas second-order moments are studied by two-dimensional tensors. Quality engineering maintains that we should do tuning only under standard conditions after completing robust design. A two-dimensional tensor does not necessarily represent noises in an SN ratio. In quality engineering, noise effects should have a continual monotonic tendency.

Quality engineering says that we should adjust a function to a target value or curve under standard conditions only after reducing variability. The idea is that, first, robust design or functional improvement for certain conditions of use needs to be completed, following which tuning (a method of adjusting a response value to a target value or curve) is done. Yet people in charge of designing a receiver sometimes insist that unless tuning is done first, output characteristics cannot be attained. In this case, what can we do? This is the reason that quality engineering requires us to acquire knowledge of measurement and analysis technologies. Several different measurement methods should be investigated.

A study to attain an objective value or function under 16 types of noise conditions by altering parameters can be regarded as one parameter design method; however, this is different from quality engineering. Indeed, it is true that quality characteristics have been studied by changing system parameters in a number of design of experiments cases, utilized broadly in the Japanese engineering field, and consequently, contributed to improving product quality. However, in terms of efficacy, these methods have not been satisfactory. We do not describe them in detail here because many publications do so. Improving objective characteristics by parameters, termed *design of experiments* (DOE), has the following two purposes: (1) to check out effects on objective characteristics and improve through tuning, and (2) to check out effects on output characteristics and conduct tolerance design. The objective of this chapter is to compare quality engineering and conventional methods.

Quality engineering has been used in the United States since the 1980s; however, it differs from the quality engineering detailed here, which is called the Taguchi method, the Taguchi paradigm, or Taguchi quality engineering in Europe and the United States. Quality engineering is robust design based on the following three procedures: (1) orthogonal array, (2) SN ratio, and (3) loss function. These three procedures are not robust design per se because they are used to evaluate technical means or products. To design robustness, we are required to understand the meaning of robust design before considering technical or management means. Those who consider practical means are engineers and production people.

Robust design (including product and process design) generally means designing a product that can function properly under various conditions of use. If this is its lone objective, we can study reliability or fraction nondefective, whose goal is 100%. Yet a corporation is merely a means of making a profit. Increased cost caused by excessive quality design leads to poor profitability and potentially to bankruptcy. The true objective of design and production is to earn a corporate profit.

The essence of the Taguchi method lies in measurement in functional space (to calculate a signal's proximity to an ideal function of output using the reciprocal of the average sum of squared deviations from an ideal function), overall measurement in multidimensional space from a base point, and accuracy. Quality engineering involves a research procedure for improving functionality and accuracy of diagnosis or prediction by taking advantage of measurement methods in functional space and in multidimensional space. Each improvement is an individual engineer's task. In the Taguchi method for hardware design, downstream reproducibility is important. For this purpose, it is desirable to collect data whose input signal and output characteristic are energy-related. Also, analysis can be conducted using the square root of the data. On the other hand, in case of diagnosis and prediction, as long as multidimensional data have sufficient accuracy, the equation of Mahalanobis distance in homogeneous or unit space is used. To measure the validity of use, an SN ratio is used by calculating the Mahalanobis distance from objects that obviously do not belong to the homogeneous space. The SN ratio is used to examine the effect of control factors. In the case of hardware study, control factors include design constants, and in the case of multidimensional data, control vectors include the selection of items. The reproducibility of control factor effects is checked using orthogonal arrays.

In this chapter we explain the Taguchi method in the field of quality engineering by explaining how to use it in several typical technological fields. Since the concepts apply for general technological fields, the Taguchi method is classified as a generic technology. In this chapter we briefly explain its application to each of the following fields: (1) electronic and electrical engineering, (2) mechanical engineering, (3) chemical engineering, (4) medical treatment and efficacy experimentation, (5) software testing, (6) MT and MTS Methods, and (7) on-line quality engineering.

Sections 4.2 to 4.5 cover quality engineering for hardware. Section 4.4 includes engine function as a generic function. In Section 4.6 we discuss debugging tests in terms of software quality. Section 4.7 highlights a new quality engineering method, which includes pattern recognition in regard to diagnosis and prediction. Section 4.8 addresses the system design of feedback quality management in the category of the on-line quality engineering. All formulas used in this chapter are given in Chapter 3 of Reference 1.

4.2. Electronic and Electrical Engineering

In general, the electronic and electrical engineering field emphasizes the following two functions: to supply power and to supply information. Power includes not only electric power but also output, such as current and voltage. However, these functions are almost the same as those in mechanical system design, as discussed in the next section. Thus, as a unique function in electronic and electrical engineering (and in mechanical engineering), we explain experimentation and evaluation of functionality in cases where both input signal and output data are expressed in complex numbers. The electronic and electrical technology regarding information systems generally consists of the following three types:

1. *Systems using power (amplitude)*: power supply or AM (amplitude modulation) system
2. *Systems using frequency:* measurement or FM (frequency modulation) system
3. *Systems using phase:* PM (phase modulation) system

After elucidating the significance of two-stage optimization, which is common to all categories, we provide details. Even a system classified as 2 or 3 above needs stability and functionality to power. In contrast, a receiving system (or transmitting and receiving system) focuses more on accuracy in receiving information than on stability of power. Therefore, we should develop a system by considering each functionality carefully. Energy difference for frequency difference is too small to measure directly. As for phase, energy difference for phase difference is so small that it cannot be measued. We need to calculate the SN ratio and sensitivity by measuring frequency or phase itself for 2 or 3. Therefore, a procedure is needed for studying functionality by computing the frequency or phase data in system 2 or 3, as explained later.

Quality engineering proposes that for all three categories above, after creating as complicated a circuit or system as possible, we implement parameter design to determine optimal levels of design parameters in the first stage and adjust them to target values in the second stage. This is called *two-stage optimization.*

An ideal function shows a relationship between a signal and an output characteristic under certain conditions of use created by engineers. Indeed, no ideal function exists in nature; however, the process of studying and designing a function so as to match it with an ideal function is research and development. In quality engineering, we conduct research by following two steps:

1. Reduce the variability of a function under various conditions.
2. Bring the function close to an ideal function under standard conditions.

The first step is evaluation by the SN ratio; the second, called *adjustment* or *tuning*, brings a function close to an ideal by using traditional methods such as least squares. Most conventional research has been conducted in reverse, that is, the second step, then the first. In 1984, we visited the Circuit Laboratory, one of the Bell Labs, for the first time. They were developing a new circuit. As a first step, they developed it such that its function as a signal could match an objective function under standard conditions. Next, they tested whether its function could work under 16 different types of conditions (environmental conditions or noise conditions, including a deterioration test). If it functioned well under the 16 conditions, its development was completed. However, if not, they altered design constants (parameters) so that the objective function would work under all 16 conditions. Now, setting design constants to A, B, ... , and the 16 noise conditions to N_1, N_2, ... , N_{16}, we obtain

$$f(A, B, \ldots, N_i) = m \qquad (i = 1, 2, \ldots, 16) \tag{4.1}$$

In this case, m is a target value (or target curve). If we have 16 or more control factors (called *adjusting factors* in quality engineering), we can solve equation (4.1). If not, we determine A, B, ... , to minimize a sum of squared differences between both sides of equation (4.1). In other words, we solve the following:

$$\min_{A,B,\ldots} \sum_{i=1}^{16} [f(A,B, \ldots, N_i) - m]^2 \tag{4.2}$$

This method is termed *least squares.* Although functional improvement based on this method has long been used, the problem is that we cannot predict how the function behaves under conditions other than the preselected 16 conditions. It should be impossible to adjust the function to be objective function under a variety of conditions of use.

Tuning or adjustment in quality engineering, adjust a certain value to a target value or curve, prescribes that tuning must be done only under standard conditions. What type of characteristic or target curve a product should have under standard conditions is a user's decision and an ideal function in design. Although a target value or curve of a product can be selected by a designer before shipping, it is predetermined by customers of the product or parent companies. In both cases, the target value or curve needs to be studied under standard conditions.

This is because a production engineering or manufacturing department should produce products with a target value or function when shipped. Even if a product meets the target function at the point of shipment, as conditions of use vary or the product itself deteriorates when used, it sometimes malfunctions. This market quality issue is regarded as a functionality problem for which a design department should take full responsibility. The main objective of quality engineering is to take measures for functional variability under various conditions of use in the market. Indeed, we can study an ideal function for signals under standard conditions; however, we need to improve functionality for noises in the market when designing a product. When there is trouble in the market, no operator of a production plant can adjust the functions in that market. Thus, a designer should not only maintain effects for signals but also minimize functional variability caused by noises.

Functionality Evaluation of System Using Power (Amplitude)

EFFECTIVE AND INEFFECTIVE ELECTRIC POWER

In quality engineering, to assess functionality, we regard a sum of squared measurements as total output, decompose it into effective and harmful parts, and improve the ratio of the effective part to the harmful part by defining it as an SN ratio. To decompose measurements properly into quadratic forms, each measurement should be proportional to the square root of energy. Of course, the theory regarding quadratic form is related not to energy but to mathematics. The reason that each measurement is expected to be proportional to a square root of energy is that if total variation is proportional to energy or work, it can be decomposed into a sum of signals or noises, and each gain can be added sequentially.

In an engineering field such as radio waves, electric power can be decomposed as follows:

$$\text{apparent power} = \text{effective power} + \text{reactive power}$$

The effective power is measured in watts, whereas the apparent power is measured in volt-amperes. For example, suppose that input is defined as the following sinusoidal voltage:

$$\text{input} = E \sin(\omega t + \alpha) \tag{4.3}$$

and output is expressed by the following current:

$$\text{output} = I \sin(\omega t + \beta) \tag{4.4}$$

In this case, the effective power W is expressed as

$$W = \frac{EI}{2} \cos(\alpha - \beta) \tag{4.5}$$

$\cos(\alpha - \beta)$ is the *power factor*. On the other hand, since the

$$\text{apparent power} = \frac{EI}{2} \tag{4.6}$$

the reactive power is expressed by

$$\text{reactive power} = \frac{EI}{2} [1 - \cos(\alpha - \beta)] \tag{4.7}$$

Since in the actual circuit (as in the mechanical engineering case of vibration), phases α and β vary, we need to decompose total variation into parts, including not only effective energy but also reactive energy. This is the reason that variation decomposition by quadratic form in complex number or positive Hermitian form is required.

OUTPUT DECOMPOSITION BY HERMITIAN FORM

We set the signal to M, measurement to y, and the proportionality between both to

$$y = \beta M \tag{4.8}$$

If we know an ideal theoretical output, y should be selected as loss. If M, y, and β are all complex numbers, how can we deal with these? When y is a real number, defining the total variation of measurement y as total output S_T, we decompose S_T into useful and harmful parts:

$$S_T = y_1^2 + y_2^2 + \cdots + y_n^2 \tag{4.9}$$

Using the Hermitian form, which deal with the quadratic form in complex number, we express total variation as

$$S_T = y_1\bar{y}_1 + y_2\bar{y}_2 + \cdots + y_n\bar{y}_n \tag{4.10}$$

Now $\bar{y}$ is a complex conjugate number of y. What is important here is that S_T in equation (4.10) is a positive (strictly speaking, nonnegative) real number which reflects not only real number parts but also imaginary number parts.

The proportionality coefficient β is estimated by

$$\beta = \frac{\overline{M}_1 y_1 + \overline{M}_2 y_2 + \cdots + \overline{M}_n y_n}{M_1\overline{M}_1 + M_2\overline{M}_2 + \cdots + M_n\overline{M}_n} \tag{4.11}$$

In addition, the complex conjugate number of β is expressed by

$$\bar{\beta} = \frac{M_1\bar{y}_1 + M_2\bar{y}_2 + \cdots + M_n\bar{y}_n}{M_1\overline{M}_1 + M_2\overline{M}_2 + \cdots + M_n\overline{M}_n} \tag{4.12}$$

The variation of proportional terms is calculated as

$$\begin{aligned} S_\beta &= \beta\bar{\beta}(M_1\overline{M}_1 + M_2\overline{M}_2 + \cdots + M_n\overline{M}_n) \\ &= \frac{(\overline{M}_1 y_1 + \cdots + \overline{M}_n y_n)(M_1\bar{y}_1 + \cdots + M_n\bar{y}_n)}{M_1\overline{M}_1 + M_2\overline{M}_2 + \cdots + M_n\overline{M}_n} \qquad (f = 1) \end{aligned} \tag{4.13}$$

Therefore, the error variation is computed as

$$S_e = S_T - S_\beta \tag{4.14}$$

Since, in this case, a compounded noise factor is not included,

$$V_N = V_e = \frac{S_e}{n - 1} \tag{4.15}$$

The SN ratio and sensitivity are calculated as

$$\eta = 10 \log \frac{(1/r)\ (S_\beta - V_e)}{V_e} \tag{4.16}$$

$$S = 10 \log \frac{1}{r}\ (S_\beta - V_e) \tag{4.17}$$

Now r (effective divider), representing a magnitude of input, is expressed as

$$r = M_1\overline{M}_1 + M_2\overline{M}_2 + \cdots + M_n\overline{M}_n \tag{4.18}$$

In cases where there is a three-level compounded noise factor N, data can be tabulated as shown in Table 4.1 when a signal M has k levels. L_1, L_2, and L_3 are linear equations computed as

$$L_1 = \overline{M}_1 y_{11} + \overline{M}_2 y_{12} + \cdots + \overline{M}_k y_{1k}$$

$$L_2 = \overline{M}_1 y_{21} + \overline{M}_2 y_{22} + \cdots + \overline{M}_k y_{2k}$$

$$L_1 = \overline{M}_1 y_{31} + \overline{M}_2 y_{32} + \cdots + \overline{M}_k y_{3k} \tag{4.19}$$

Based on these conditions, we can find the variation and variance as follows:

Total variation:

$$S_T = y_{11}\overline{y}_{11} + y_{12}\overline{y}_{12} + \cdots + y_{3k}\overline{y}_{3k} \qquad (f = 3k) \tag{4.20}$$

Variation of proportional terms:

$$S_\beta = \frac{(L_1 + L_2 + L_3)(\overline{L}_1 + \overline{L}_2 + \overline{L}_3)}{3r} \qquad (f = 1) \tag{4.21}$$

Effective divider:

$$r = M_1\overline{M}_1 + M_2\overline{M}_2 + \cdots + M_k\overline{M}_k \tag{4.22}$$

Variation of sensitivity variability:

$$S_{N\beta} = \frac{L_1\overline{L}_1 + L_2\overline{L}_2 + L_3\overline{L}_3}{r} - S_\beta \qquad (f = 2) \tag{4.23}$$

Error variation:

$$S_e = S_T - S_\beta - S_{N\beta} \qquad (f = 3k - 3) \tag{4.24}$$

Table 4.1
Input/output data

	Signal				
Noise	M_1	M_2	···	M_k	**Linear Equation**
N_1	y_{11}	y_{12}	···	y_{1k}	L_1
N_2	y_{21}	y_{22}	···	y_{2k}	L_2
N_3	y_{31}	y_{32}	···	y_{3k}	L_3

Total noise variation:

$$S_N = S_e + S_{N\beta} \qquad (f = 3k - 1) \tag{4.25}$$

Error variance for correction:

$$V_e = \frac{S_e}{3k - 3} \tag{4.26}$$

Noise magnitude including nonlinearity (total error variance):

$$V_N = \frac{S_N}{3k - 1} \tag{4.27}$$

SN ratio:

$$\eta = 10 \log \frac{(1/3r)\ (S_\beta - V_e)}{V_N} \tag{4.28}$$

Sensitivity:

$$S = 10 \log \frac{1}{3r} (S_\beta - V_e) \tag{4.29}$$

We have thus far described application of the Hermitian form to basic output decomposition. This procedure is applicable to decomposition of almost all quadratic forms because all squared terms are replaced by products of conjugates when we deal with complex numbers. From here on we call it by a simpler term, *decomposition of variation,* even when we decompose variation using the Hermitian form because the term *decomposition of variation by the Hermitian form* is too lengthy.

Moreover, we wish to express even a product of two conjugates y and $\bar{y}$ as y^2 in some cases. Whereas y^2 denotes itself in the case of real numbers, it denotes a product of conjugates in the case of complex numbers. By doing this, we can use both the formulas and degrees of freedom expressed in the positive quadratic form. Additionally, for decomposition of variation in the Hermitian form, decomposition formulas in the quadratic form hold true. However, as for noises, we need to take into account phase variability as well as power variability because the noises exist in the complex plane. Thus, we recommend that the following four-level noise N be considered:

N_1: negative condition for both power and phase

N_2: negative condition for power, positive for phase

N_3: positive condition for power, negative for phase

N_4: positive condition for both power and phase

Change of phase is equivalent to change of reactance. By studying noises altering reactance, we set up positive and negative levels.

Except for noise levels, we can calculate the SN ratio in the same manner as that for the quadratic form. After maximizing the SN ratio, we aim at matching sensitivity β (a complex number) with a target value. To adjust both real and imaginary parts to target values, instead of sensitivity β^2 we need to consider control factor effects for β itself.

If we compute a theoretical output value in case of no loss, by redefining a deviation from a theoretical output value as an output datum y, we analyze with SN ratio and sensitivity. Although the calculation procedure is the same, sensitivity with loss should be smaller. In some cases, for a smaller-the-better function, we use a smaller-the-better SN ratio using all data.

Quality Engineering of System Using Frequency

HOW TO HANDLE FREQUENCY

Setting frequency to f and using Planck's constant h, we obtain frequency energy E as follows:

$$E = hf \tag{4.30}$$

When E is expressed in joules (J), h becomes the following tiny number:

$$h = 6.62 \times 10^{-34} \text{ J/s} \tag{4.31}$$

In this case, E is too small to measure. Since we cannot measure frequency as part of wave power, we need to measure frequency itself as quantity proportional to energy. As for output, we should keep power in an oscillation system sufficiently stable. Since the system's functionality is discussed in the preceding section, we describe only a procedure for measuring the functionality based on frequency.

If a signal has one level and only stability of frequency is in question, measurement of time and distance is regarded as essential. Stability for this case is nominal-the-best stability of frequency, which is used for measuring time and distance. More exactly, we sometimes take a square root of frequency.

STABILITY OF TRANSMITTING WAVE

The stability of the transmitting wave is considered most important when we deal with radio waves. (A laser beam is an electromagnetic wave; however, we use the term *radio wave*. The procedure discussed here applies also for infrared light, visible light, ultraviolet light, x-rays, and gamma rays.) The stability of radio waves is categorized as either stability of phase or frequency or as stability of power. In the case of frequency-modulated (FM) waves, the former is more important because even if output fluctuates to some degree, we can easily receive phase or frequency properly when phase or frequency is used.

In Section 7.2 of Reference 2, for experimentation purposes the variability in power source voltage is treated as a noise. The stability of the transmitting wave is regarded as essential for any function using phase or frequency. Since only a single frequency is sufficient to stabilize phase in studying the stability of the transmitting wave, we can utilize a nominal-the-best SN ratio in quality engineering. More specifically, by selecting variability in temperature or power source voltage or deterioration as one or more noises, we calculate the nominal-the-best SN ratio. If the noise has four levels, by setting the frequency data at N_1, N_2, N_3, and N_4 to y_1, y_2, y_3, and y_4, we compute the nominal-the-best SN ratio and sensitivity as follows:

$$\eta = 10 \log \frac{\frac{1}{4}(S_m - V_e)}{V_e} \tag{4.32}$$

$$S = 10 \log \tfrac{1}{4}(S_m - V_e) \tag{4.33}$$

Now

$$S_m = \frac{(y_1 + y_2 + y_3 + y_4)^2}{4} \qquad (f = 1) \tag{4.34}$$

$$S_T = y_1^2 + y_2^2 + y_3^2 + y_4^2 \qquad (f = 4) \tag{4.35}$$

$$S_e = S_T - S_m \qquad (f = 3) \tag{4.36}$$

$$V_e = \tfrac{1}{3}S_e \tag{4.37}$$

After determining the optimal SN ratio conditions, we perform a confirmatory experiment under those conditions and compute the corresponding frequency. When dealing with a clock, its frequency can be set to 1 second. If we have a target value for a frequency, we can bring the frequency close to the target by using control factors affecting sensitivity S greatly or changing the effective inductance or capacitance. Since tuning is relatively easy for radio waves, sensitivity data are often of no importance.

To generalize this procedure, we recommend using multiple frequencies. We select K levels from a certain band and measure data at two compounded noises, N_1 and N_2, as illustrated in Table 4.2. For signal M, k levels can be selected arbitrarily in a certain band. For example, when we obtain k levels by changing the inductance, we do not need to consider an ideal relationship between inductance and M. If an equation representing such a relationship exists, it does not matter so much, and we do not have to include a deviation from the equation in errors. Therefore, we can decompose total variation as shown in Table 4.3. The effect of signal M is the effect of all k levels.

$$S_M = \frac{(y_{11} + y_{12})^2 + \cdots + (y_{k1} + y_{k2})^2}{2} \qquad (f = k) \tag{4.38}$$

$$S_N = \frac{[(y_{11} + \cdots + y_{k1}) - (y_{12} + \cdots + y_{k2})]^2}{2k} \tag{4.39}$$

$$S_e = S_T - S_M - S_N \tag{4.40}$$

S_T is now the total variation of output.

Table 4.2
Cases where signal factor exists

	Noise		
Signal	**N_1**	**N_2**	**Total**
M_1	y_{11}	y_{12}	y_1
M_2	y_{21}	y_{22}	y_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
M_k	y_{k1}	y_{k2}	y_k

Table 4.3
Decomposition of total variation

Factor	f	S	V
M	k	S_M	V_M
N	1	S_N	
e	$k - 1$	S_e	V_e
Total	$2k$	S_T	

The SN ratio and sensitivity are as follows:

$$\eta = 10 \log \frac{\frac{1}{2}(V_M - V_e)}{V_N} \tag{4.41}$$

$$S = 10 \log \tfrac{1}{2}(V_M - V_e) \tag{4.42}$$

Now

$$V_M = \frac{S_M}{k} \tag{4.43}$$

$$V_e = \frac{S_e}{k - 1} \tag{4.44}$$

$$V_N = \frac{S_N + S_e}{k} \tag{4.45}$$

Since we consider stabilizing all frequencies in a certain range of frequency, this procedure should be more effective than a procedure of studying each frequency.

THE CASE OF MODULATION FUNCTION

When we transmit (transact) information using a certain frequency, we need frequency modulation. The technical means of modulating frequency is considered to be the signal. A typical function is a circuit that oscillates a radio wave into which we modulate a transmitting wave at the same frequency as that of the transmitting wave and extract a modulated signal using the difference between two waves. By taking into account various types of technical means, we should improve the following proportionality between input signal M and measurement characteristic y for modulation:

$$y = \beta M \qquad (y \text{ and } M \text{ take both positive and negative numbers}) \tag{4.46}$$

Thus, for modulation functions, we can define the same SN ratio as before. Although y as output is important, before studying y we should research the frequency stability of the transmitting wave and internal oscillation as discussed in the preceding section.

When we study the relationship expressed in equation (4.46) by setting a difference between a transmitting wave and an internally oscillated radio wave to output, it is necessary for us to measure data by changing not only conditions regarding the environment or deterioration as noises but also frequency (not a modulated signal but a transmitting wave). If we select two levels for noise N, three levels for frequency F, and k levels for modulated signal M, we obtain the data shown in Table 4.4. Frequency F is called an *indicative factor*. According to this, we decompose total variation as shown in Table 4.5.

Now r represents the total variation of the input modulated signal. The calculation procedure is as follows:

$$S_\beta = \frac{(L_1 + L_2 + \cdots + L_6)^2}{6r} \qquad (f = 1) \tag{4.47}$$

$$S_{F\beta} = \frac{(L_1 + L_4)^2 + \cdots + (L_3 + L_6)^2}{2r} - S_\beta \qquad (f = 2) \tag{4.48}$$

$$S_{N(F)\beta} = \frac{L_1^2 + \cdots + L_6^2}{r} - S_\beta - S_{F\beta} \qquad (f = 3) \tag{4.49}$$

$$S_e = S_T - (S_\beta + S_{F\beta} + S_{N(F)\beta}) \qquad (F = 6k - 6) \tag{4.50}$$

Using these results, we calculate the SN ratio and sensitivity:

$$\eta = 10 \log \frac{(1/6r)(S_\beta - V_e)}{V_N} \tag{4.51}$$

$$S = 10 \log \frac{1}{6r}(S_\beta - V_e) \tag{4.52}$$

Now

$$V_N = \frac{S_{N(F)\beta} + S_e}{6k - 3} \tag{4.53}$$

FUNCTIONALITY OF SYSTEM USING PHASE

In the case of analog modulation, modulation technology has evolved from AM and FM to PM. Indeed, digital systems also exist; however, in quality engineering

Table 4.4
Case where modulated signal exists

Noise	Frequency	Modulated Signal M_1	M_2	⋯	M_k	Linear Equation
N_1	F_1	y_{11}	y_{12}	⋯	y_{1k}	L_1
	F_2	y_{21}	y_{22}	⋯	y_{2k}	L_2
	F_3	y_{31}	y_{32}	⋯	y_{3k}	L_3
N_2	F_1	y_{41}	y_{42}	⋯	y_{4k}	L_4
	F_2	y_{51}	y_{52}	⋯	y_{5k}	L_5
	F_3	y_{61}	y_{62}	⋯	y_{6k}	L_6

Table 4.5
Decomposition of total variation

Factor	f	S	V
β	1	S	
$F\beta$	2	$S_{F\beta}$	
$N(F)\beta$	3	$S_{N(F)\beta}$	
e	$6(k-1)$	S_e	V_e
Total	$6k$	S_T	

we should research and design only analog functions because digital systems are included in all AM, FM, and PM, and its only difference from analog systems is that its level value is not continuous but discrete: that is, if we can improve SN ratio as an analog function and can minimize each interval of discrete values and eventually enhance information density.

❒ Example

A phase-modulation digital system for a signal having a 30° phase interval such as 0, 30, 60, ... , 330° has a σ value that is expressed in the following equation, even if its analog functional SN ratio is −10 dB:

$$10 \log \frac{\beta^2}{\sigma^2} = -10 \tag{4.54}$$

By taking into account that the unit of signal M is degrees, we obtain the following small value of σ:

$$\sigma = 31.6\beta = 3.16° \tag{4.55}$$

The fact that σ representing RMS is approximately 3° implies that there is almost no error as a digital system because 3° represents one-fifth of a function limit regarded as an error in a digital system. In Table 4.6 all data are expressed in radians. Since the SN is 48.02 dB, σ is computed as follows:

$$\begin{aligned} \sigma &= \sqrt{10^{-4.802}}\ \beta = 0.00397 \text{ rad} \\ &= 0.22° \end{aligned} \tag{4.56}$$

The ratio of this σ to the function limit of $\pm 15°$ is 68. Then, if noises are selected properly, almost no error occurs.

In considering the phase modulation function, in most cases frequency is handled as an indicative factor (i.e., a use-related factor whose effect is not regarded as noise.) The procedure of calculating its SN ratio is the same as that for a system

Table 4.6
Data for phase shifter (experiment 1 in L_{18} orthogonal array; rad)

		Voltage		
Temperature	**Frequency**	V_1	V_2	V_3
T_1	F_1	$77 + j101$	$248 + j322$	$769 + j1017$
	F_2	$87 + j104$	$280 + j330$	$870 + j1044$
	F_3	$97 + j106$	$311 + j335$	$970 + j1058$
T_2	F_1	$77 + j102$	$247 + j322$	$784 + j1025$
	F_2	$88 + j105$	$280 + j331$	$889 + j1052$
	F_3	$98 + j107$	$311 + j335$	$989 + j1068$

using frequency. Therefore, we explain the calculation procedure using experimental data in *Technology Development for Electronic and Electric Industries* [3].

❒ Example

This example deals with the stability design of a phase shifter to advance the phase by 45°. Four control factors are chosen and assigned to an L_{18} orthogonal array. Since the voltage output is taken into consideration, input voltages are selected as a three-level signal. However, the voltage V is also a noise for phase modulation. Additionally, as noises, temperature T and frequency F are chosen as follows:

Voltage (mV): $V_1 = 224$, $V_2 = 707$, $V_3 = 2234$

Temperature (°C): $T_1 = 10$, $T_2 = 60$

Frequency (MHz): $F_1 = 0.9$, $F_2 = 1.0$, $F_3 = 1.1$

Table 4.6 includes output voltage data expressed in complex numbers for 18 combinations of all levels of noises V, T, and F. Although the input voltage is a signal for the output voltage, we calculate the phase advance angle based on complex numbers because we focus only on the phase-shift function:

$$y = \tan^{-1} \frac{Y}{X} \tag{4.57}$$

In this equation, X and Y indicate real and imaginary parts, respectively. Using a scalar Y, we calculate the SN ratio and sensitivity.

As for power, we compute the SN ratio and sensitivity according to the Hermitian form described earlier in the chapter. Because the function is a phase advance (only 45° in this case) in this section, the signal has one level (more strictly, two phases, 0 and 45°). However, since our focal point is only the difference, all that we need to do is to check phase advance angles for the data in Table 4.6. For the sake of convenience, we show them in radians in Table 4.7, whereas we can also use degrees as the unit. If three phase advance angles, 45, 90, and 135°, are selected, we should consider dynamic function by setting the signal as M.

Table 4.7
Data for phase advance angle (rad)

Temperature	Frequency	Voltage V_1	V_2	V_3	Total
T_1	F_1	0.919	0.915	0.923	2.757
	F_2	0.874	0.867	0.876	2.617
	F_3	0.830	0.823	0.829	2.482
T_2	F_1	0.924	0.916	0.918	2.758
	F_2	0.873	0.869	0.869	2.611
	F_3	0.829	0.823	0.824	2.476

Since we can adjust the phase angle to 45° or 0.785 rad (the target angle), here we pay attention to the nominal-the-best SN ratio to minimize variability. Now T and V are true noises and F is an indicative factor. Then we analyze as follows:

$$S_T = 0.919^2 + 0.915^2 + \cdots + 0.824^2 = 13.721679$$

$$(f = 18) \tag{4.58}$$

$$S_m \frac{15.701^2}{18} = 13.695633$$

$$(f = 1) \tag{4.59}$$

$$S_F = \frac{5.515^2 + 5.228^2 + 4.958^2}{6} - S_m = 0.0258625$$

$$(f = 2) \tag{4.60}$$

$$S_e = S_T - S_m - S_F = 0.0001835$$

$$(f = 15) \tag{4.61}$$

$$V_e = \frac{S_e}{15}\,(0.00001223) \tag{4.62}$$

Based on these results, we obtain the SN ratio and sensitivity as follows:

$$\eta = 10 \log \frac{\frac{1}{18}(13.695633 - 0.00001223)}{0.00001223}$$

$$= 47.94 \text{ dB} \tag{4.63}$$

$$S = 10 \log \tfrac{1}{18}(13.695633 - 0.00001223)$$

$$= -1.187 \text{ dB} \tag{4.64}$$

The target angle of 45° for signal S can be expressed in radians as follows:

$$S = 10 \log \frac{45}{180} (3.1416) = -1.049 \text{ dB} \tag{4.65}$$

As a next step, under optimal conditions, we calculate the phase advance angle and adjust it to the value in equation (4.65) under standard conditions. This process can be completed by only one variable.

When there are multiple levels for one signal factor, we should calculate the dynamic SN ratio. When we design an oscillator using signal factors, we may select three levels for a signal to change phase, three levels for a signal to change output, three levels for frequency F, and three levels for temperature as an external condition, and assign them to an L_9 orthogonal array. In this case, for the effective value of output amplitude, we set a proportional equation for the signal of voltage V as the ideal function. All other factors are noises. As for the phase modulation function, if we select the following as levels of a phase advance and retard signal,

$$\textit{Level targeted at } -45°\text{:} \quad M_1' - M_2' = m_1$$

$$\textit{Level targeted at } 0°\text{:} \quad M_2' - M_2' = m_2$$

$$\textit{Level targeted at } +45°\text{:} \quad M_3' - M_2' = m_3$$

the signal becomes zero-point proportional. The ideal function is a proportional equation where both the signal and output go through the zero point and take positive and negative values. If we wish to select more signal levels, for example, two levels for each of the positive and negative sides (in total, five levels, M_1, M_2, M_3, M_4, and M_5), we should set them as follows:

$$M_1 = -120°$$

$$M_2 = -60°$$

$$M_3 = M_4 = 0 \quad \text{(dummy)}$$

$$M_5 = +\ 60°$$

$$M_6 = +\ 120°$$

Next, together with voltage V, temperature T, and frequency F, we allocate them to an L_{18} orthogonal array to take measurements. On the other hand, by setting $M_1 = -150°$, $M_2 = -90°$, $M_3 = -30°$, $M_4 = +30°$, $M_5 = +90°$, and $M_6 = +150°$, we can assign them to an L_{18} array. After calibrating the data measured using $M_0 = 0°$, they are zero proportional.

DESIGN BY SIMULATION

In electronic and electric systems, theories showing input/output relationships are available for analysis. In some cases, approximate theories are available. Instead of experimenting, we can study functionality using simulation techniques on the computer. About half of the electrical and electronic design examples in References 1 and 4 are related to functionality (reliability) design by simulation, even though there are not as many electrical and electronic examples. Even if a theory is imperfect or partial, functionality design often goes well. This is because in quality engineering we study basic functionality first and then adjust the outputs to target values with a few parameters on an actual circuit or machine.

Chapter 16 of Reference 4 illustrates Hewlett-Packard's simulation design using complex numbers, in which real and imaginary parts are dealt with separately for parameter design. Ideally, it would be better to design parameters for power according to the procedure discussed earlier and to study phase stability based on the technique shown earlier. It was necessary to allocate all control factors selected to an L_{108} orthogonal array. In place of analyzing real and imaginary parts separately, parameter design should be performed using the data of effective amplitude and phase angle. However, decomposition of total variation by the Hermitian form was not known at that time.

If simulation calculation does not take long, after creating three noise factor levels before and after each control factor level, we should assign them to an orthogonal array. Then, to compute the SN ratio, we measure data for the noise-assigned orthogonal array (called the *outer array*) corresponding to each experiment in the orthogonal array.

In the twenty-first century, simulation will play a major role in R&D. In this situation, partially established theories or approximation methods suffice for studying functional variability because such techniques can help improve robustness. Tuning the outputs to the objective functions may be performed using hardware.

4.3. Mechanical Engineering

Conventional Meaning of Robust Design

A research way of reducing variability only because an objective function is varying is regarded as *whack-a-mole*, or unplanned. Although this problem-solving method should not be used, engineers have been trying to prevent problems with a new product caused by current research and design. Ideally, we should prevent quality problems in the manufacturing process or the market through optimization at the design stage.

Although the Nippon Telegraph and Telephone Public Corporation (NTT) designed exchangers or telephones in the 1950s, those who produced these products were private companies such as NEC, Hitachi, Fujitsu, and Oki. Therefore, without knowing how to produce them, NTT designed them in such a way that they could prevent quality problems in the market.

While the system realization department conceptualized future transmission systems and requested devices, units, materials, and parts with new functions needed at that time, the device realization department developed new products and technologies to meet the requests. That is, the telecommunication laboratory did not have a design department. The results from research and development were drawings and specifications, and the manufacturers of communication devices noted above produced products according to these specs.

A design department is responsible for avoiding problems in the manufacturing process or the marketplace before mass production or shipping is begun. Therefore, in the design phase, manufacturers had to improve quality (functional variability) in the manufacturing process or the marketplace without knowing the type of manufacturing process that would be used or how quality would be controlled. In fact, no problems occurred. Researchers knew the objective functions of products they had developed themselves. In fact, they knew not only the objective functions but also the generic functions that were the means to attaining the objective functions. The mistake that researchers in Europe and the United States made in evaluating objective or generic functions was not related to the creation of technical means.

Although these researchers are creative and talented, their research procedure was first to develop and design a function under standard conditions, then test functionality under those conditions and study countermeasures against the problems through design changes. Then, by modifying product design, they take countermeasures for functional variability (functionality) caused by conditions of production or use after discovering functional problems through tests under various conditions. Needless to say, this can be regarded as "whack-a-mole" research.

Quality engineering recommends that we modify product design using an L_{18} orthogonal array before finding problems. Researchers quite often say that they cannot consider technical measures without looking at actual problems. The conventional design procedure comprises the following steps:

1. Under standard conditions, we design a product or system that each of us considers optimal.
2. To find design problems, we test the product or system under various conditions.
3. We change the design of the product or system (called *tuning*, a countermeasure based on cause-and-effect relationships) such that it functions well even under conditions where the original design does not work properly.

Quite often the second and third steps are repeated. We call this procedure "whack-a-mole." Quality engineering insists that such research methods be stopped.

New Method: Functionality Design

Quality engineering recommends improving functionality before a quality problem (functional variability) occurs. In other words, before a problem happens, a designer makes improvements. Quite a few researchers or designers face difficulties understanding how to change their designs. Quality engineering advises that we check on changes in functionality using SN ratios by taking advantage of any type of design change. A factor in design change is called a *control factor*. Most control factors should be assigned to an L_{18} orthogonal array. A researcher or designer should have realized and defined both the objective and generic functions of the product being designed.

❒ Example

Imagine a vehicle steering function, more specifically, when we change the orientation of a car by changing the steering angle. The first engineer who designed such a system considered how much the angle could be changed within a certain range of steering angle. For example, the steering angle range is defined by the rotation of a steering wheel, three rotations for each, clockwise and counterclockwise, as follows:

$$(-360)(3) = -1080°$$

$$(360)(3) = +1080°$$

In this case, the total steering angle adds up to 2160°. Accordingly, the engineer should have considered the relationship between the angle above and steering curvature y, such as a certain curvature at a steering angle of $-1080°$, zero curvature at 0, and a certain curvature at $+1080$.

For any function, a developer of a function pursues an ideal relationship between signal M and output y under conditions of use. In quality engineering, since any function can be expressed by an input/output relationship of work or energy, essentially an ideal function is regarded as a proportionality. For instance, in the steering function, the ideal function can be given by

$$y = \beta M \tag{4.66}$$

In this equation, y is curvature (the reciprocal of a turning radius) and M is a steering angle range. European or U.S. researchers (especially inventors of functions) who develop a product or system know that a relationship exists between its signal and output characteristic y. They make efforts to design functions and invent a number of techniques in order to match a function with an ideal function under standard conditions. This matching procedure known as *tuning*, utilizes the least-squares method in mathematical evaluation.

Quality engineering maintains that we first improve functionality (functional variability) under conditions of use by deferring until later the adjustment of a function to the ideal function under standard conditions. Conditions of use are various environmental conditions under which a product or system is used, including both environmental conditions and deterioration. For the case of the steering function, some environmental conditions are road conditions and a tire's frictional condition, both of which are called *noises*. In quality engineering we compound all noises into only two levels, N_1 and N_2, no matter how many there are.

Now, we consider compounding noises for the steering function. For the output, curvature of radius, we define two levels of a compounded error factor N:

N_1: conditions where the proportionality coefficient β becomes smaller

N_2: Conditions where the proportionality coefficient β becomes larger

In short, N_1 represents conditions where the steering angle functions well, whereas N_2 represents conditions where it does not function well. For example, the former represents a car with new tires running on a dry asphalt road. In contrast, the latter is a car with worn tires running on a wet asphalt or snowy road.

By changing the vehicle design, we hope to mitigate the difference between N_1 and N_2. Quality engineering actively takes advantage of the interaction between control factors and a compounded error factor N to improve functionality. Then we measure data such as curvature, as shown in Table 4.8 (in the counterclockwise turn case, negative signs are added, and vice versa). Each M_i value means angle.

Table 4.8 shows data of curvature y, or the reciprocal of the turning radius for each signal value of a steering angle ranging from zero to almost maximum for both clockwise and counterclockwise directions. If we turn left or right on a normal road or steer on a highway, the range of data should be different. The data in Table 4.8 represent the situation whether we are driving a car in a parking lot or steering in a hairpin curve at low speed. The following factor of speed K, indicating different conditions of use, called an *indicative factor*, is often assigned to an outer orthogonal array. This is because a signal factor value varies in accordance with each indicative factor level.

K_1: low speed (less than 20 km/h)

K_2: middle speed (20 to 60 km/h)

K_3: high speed (more than 60 km/h)

The SN ratio and sensitivity for the data in Table 4.8 are computed as follows. First, we calculate two linear equations, L_1 and L_2, using N_1 and N_2:

$$L = M_1 y_{11} + M_2 y_{12} + \cdots + M_7 y_{17} \tag{4.67}$$

$$L_2 = M_1 y_{21} + M_2 y_{21} + M_2 y_{22} + \cdots + M_7 y_{27} \tag{4.68}$$

Next, we decompose the total variation of data in Table 4.8.

$$S_T = \text{total variation} \qquad (f = 14) \tag{4.69}$$

Variation of proportional terms:

$$S_\beta = \frac{(L_1 + L_2)^2}{2r} \tag{4.70}$$

where

$$r = M_1^2 + M_2^2 + \cdots + M_7^2$$

$$= (-720)^2 + (-480)^2 + \cdots + 720^2$$

$$= 1{,}612{,}800 \tag{4.71}$$

Table 4.8
Curvature data

	M_1 −720	M_2 −480	M_3 −240	M_4 0	M_5 240	M_6 480	M_7 720
N_1	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}	y_{17}
N_2	y_{21}	y_{22}	y_{23}	y_{24}	y_{25}	y_{26}	y_{27}

Although we choose 1° as a unit of magnitude of signal r, we can choose 120° and eventually obtain $r = 112$. That is, designers can define a unit of signal as they wish. On the other hand, the difference of proportionality coefficient for signal M (difference between linear equations N_1 and N_2), $S_{N\beta}$, is calculated

$$S_{N\beta} = \frac{(L_1 - L_2)^2}{2r} \tag{4.72}$$

Therefore, the error variation

$$S_e = S_T - (S_\beta + S_{N\beta}) \tag{4.73}$$

As a result of decomposing total variation, we obtain Table 4.9.

Finally, we can compute SN ratio and sensitivity as follows:

$$\eta = 10 \log \frac{(1/2r)\,(S_\beta - V_e)}{V_N} \tag{4.74}$$

$$\hat{\beta} = S = \frac{L_1 + L_2}{2r} \tag{4.75}$$

$$V_e = \frac{S_e}{12} \tag{4.76}$$

$$V_N = \frac{S_{N\beta} + S_e}{13} \tag{4.77}$$

The reason that sensitivity S is calculated by equation (4.75) instead of using S_β is because S_β always takes a positive value. But sensitivity is used to adjust β to target, and β may be either positive or negative, since the signal (M), takes both positive and negative values.

A major aspect in which quality engineering differs from a conventional research and design procedure is that QE focuses on the interaction between signal and noise factors, which represent conditions of design and use. As we have said

Table 4.9
Decomposition of total variation

Source	f	S	V
β	1	S_β	
$N\beta$	1	$S_{N\beta}$	$\left.\begin{matrix} \\ V_e \end{matrix}\right\} V_N$
e	12	S_e	
Total	14	S_T	

previously, in quality engineering, as a first step, we improve functionality, and second, do tuning (adjustment) on a difference between a function at an optimal SN ratio condition and an ideal function under standard conditions.

ADVANTAGES AND DISADVANTAGES OF THE NEW METHOD

Since our new method focuses on functions, we do not need to find problems through a number of tests, and setting up noise conditions does not have to be perfect. The new method has five major advantages:

1. Before product planning, it is possible to study subsystems (including overlooked items) or total system functionality. In other words, the study of functional variability at the product design stage is unnecessary, leading to earlier delivery.
2. When we improve functionality using generic functions, such developed common technologies are applicable to a variety of products. This helps not only the current project plan, but the next and consequently streamlines the research process.
3. Instead of considering all possible noise conditions such as in the traditional reliability study, only a few noise conditions selected from conditions of use are needed for analysis.
4. Since only the type of functionality is sufficient for each signal factor type, we can attain optimal design simply and quickly. Even if there are multiple types of signal factors, we can study them at the same time.
5. When systems are going to be connected, only timing is needed, without the study on variability.

The disadvantages of our new method are as follows:

1. It is necessary to persuade engineers who have never thought about the concept of signal factors to acknowledging the importance of selecting them. However, this might also be regarded as an advantage or key role of our new method because we can teach them engineering basics. That may stop engineers using easy ways to measure something related to an objective and let them look back to the essence of engineering. Some engineers work extremely hard and long hours but still often experience unexpected problems in the marketplace. The necessity of preventing such problems forces engineers to consider signal factors and to evaluate functionality using SN ratios under a variety of noise conditions because this method of improving design is quite effective.
2. Engineers need to calculate SN ratios, but this does not mean that they must do the actual calculation. Calculation is not a burden to engineers, since computer software is available. What they must do is to think about the engineering meaning of an SN ratio, then determine what type of output generated by the signal factor level must be collected. Engineers must understand the new evaluation method: Use only one SN ratio to express a function.
3. In quality engineering, when we consider a generic function, both input signal and output characteristic are not objective related but are work and energy related. However, since we have difficulties measuring work and energy in many cases, we have no other choice but to select an objective func-

tion. For example, the steering function discussed earlier is not a generic but an objective function. When an objective function is chosen, we need to check whether there is an interaction between control factors after allocating them to an orthogonal array, because their effects (differences in SN ratio, called *gain*) do not necessarily have additivity. An L_{18} orthogonal array is recommended because its size is appropriate enough to check on the additivity of gain. If there is no additivity, the signal and measurement characteristics we used need to be reconsidered for change.

4. Another disadvantage is the paradigm shift for orthogonal array's role. In quality engineering, control factors are assigned to an orthogonal array. This is not because we need to calculate the main effects of control factors for the measurement characteristic y. Since levels of control factors are fixed, there is no need to measure their effects for y using orthogonal arrays. The real objective of using orthogonal arrays is to check the reproducibility of gains, as described in disadvantage 3.

In new product development, there is no existing engineering knowledge most of the time. It is said that engineers are suffering from being unable to find ways for SN ratio improvement. In fact, the use of orthogonal arrays enables engineers to perform R&D for a totally new area. In the application, parameter-level intervals may be increased and many parameters may be studied together. If the system selected is not a good one, there will be little SN ratio improvement. Often, one orthogonal array experiment can tell us the limitation on SN ratios. This is an advantage of using an orthogonal array rather than a disadvantage.

Problem Solving and Quality Engineering

We are asked repeatedly how we would solve a problem using quality engineering. Here are some generalities that we use.

1. Even if the problem is related to a consumer's complaint in the marketplace or manufacturing variability, in lieu of questioning its root cause, we would suggest changing the design and using a generic function. This is a technical approach to problem solving through redesign.
2. To solve complaints in the marketplace, after finding parts sensitive to environmental changes or deterioration, we would replace them with robust parts, even if it incurs a cost increase. This is called *tolerance design*.
3. If variability occurs before shipping, we would reduce it by reviewing process control procedures and inspection standards. In quality engineering this is termed *on-line quality engineering*. This does not involve using Shewhart control charts or other management devices but does include the following daily routine activities by operators:
   a. *Feedback control.* Stabilize processes continuously for products to be produced later by inspecting product characteristics and correcting the difference between them and target values.
   b. *Feedforward control.* Based on the information from incoming materials or component parts, predict the product characteristics in the next process and calibrate processes continuously to match product characteristics with the target.
   c. *Preventive maintenance.* Change or repair tools periodically with or without checkup.

d. *Inspection design*. Prevent defective products from causing quality problems in downstream processes through inspection.

These are daily routine jobs designed to maintain quality level. Rational design of these systems is related to on-line quality engineering and often leads to considerable improvement.

Signal and Output in Mechanical Engineering

In terms of the generic function representing mechanical functionality,

$$y = \beta M \tag{4.78}$$

both signal M and output characteristic y should be proportional to the square root of energy or work. That is, we need to analyze using its square root because total output can be expressed as follows:

$$S_T = y_1^2 + \cdots + y_n^2 \tag{4.79}$$

By doing so, we can make S_T equivalent to work, and as a result, the following simple decomposition can be completed:

$$\text{total work} = (\text{work by signal}) + (\text{work by noise}) \tag{4.80}$$

This type of decomposition has long been used in the telecommunication field. In cases of a vehicle's acceleration performance, M should be the square root of distance and y the time to reach. After choosing two noise levels, we calculate the SN ratio and sensitivity using equation (4.80), which holds true for a case where energy or work is expressed as a square root. But in many cases it is not clear if it is so. To make sure that (4.80) holds true, it is checked by the additivity of control factor effects using an orthogonal array. According to Newton's law, the following formula shows that distance y is proportional to squared time t^2 under constant acceleration:

$$y = \frac{b}{2} t^2 \tag{4.81}$$

From a quality engineering viewpoint, it is not appropriate to use (4.81). Instead, taking ease of calculation into account, we should use

$$t = \sqrt{\frac{2}{b}} \sqrt{y}$$

$$t = \beta M \qquad \beta = \sqrt{\frac{2}{b}} \qquad \sqrt{y} = M \tag{4.82}$$

Generic Function of Machining

The term *generic function* is used quite often in quality engineering. This is not an objective function but a function as the means to achieve the objective. Since a function means to make something work, work should be measured as output. For example, in case of cutting, it is the amount of material cut. On the other hand, as work input, electricity consumption or fuel consumption can be selected. When

Ford Motor Co. experimented on functionality improvement of an engine during idling in 1995, they chose a signal, output, and ideal function as follows:

Signal M: amount of fuel flow

Output y: indicated mean effective pressure (IMEP)

Ideal function: $y = \beta M$

Now, output y indicates average pressure to rotate a crankshaft, which is regarded as a force for work. However, the generic function of an engine is a chemical reaction, discussed in the next section.

Quality engineering maintains that what is picked up as signal, what is measured, and what types of noises are chosen depends on the theme selected by researchers. Since the output of work is a request from users (consumers or operators in manufacturing), work should be selected as signal, and the amount needed for the work (in case of machinery, used power) should be measured as data. In the experiment on cutting conducted by JR in 1959 (known as National Railway at that time), the power needed for cutting was measured as net cutting power in the unit kilowatthours.

In an actual experiment, we should take measurements as follows: Selecting test pieces and cutting them for different durations, such as $T_1 = 10$ s, $T_2 = 20$ s, and $T_3 = 30$ s, we measure the amount of cut M (grams) and electricity consumption y, as shown in Table 4.10. N indicates a two-level noise. According to what these data show, we perform two different analyses.

1. *Analysis of cutting performance.* This analysis is based on the following ideal function:

$$y = \beta M \tag{4.83}$$

Since in this case, both signal of amount of cut M and data of electricity consumption y are observations, M and y have six levels and the effects of N are not considered. For a practical situation, we should take the square root of both y and M. The linear equation is expressed as

$$L = M_{11}y_{11} + M_{12}y_{12} + \cdots + M_{23}y_{23} \tag{4.84}$$

$$r = M_{11}^2 + M_{12}^2 + \cdots + M_{23}^2 \tag{4.85}$$

Table 4.10
Data for machining

	T_1	T_2	T_3
N_1	M_{11} y_{11}	M_{12} y_{12}	M_{13} y_{13}
N_2	M_{21} y_{21}	M_{22} y_{22}	M_{23} y_{23}

Regarding each variation, we find that

$$S_T = y_{11}^2 + y_{12}^2 + \cdots + y_{23}^2 \qquad (f = 6) \tag{4.86}$$

$$S_\beta = \frac{L^2}{r} \qquad (f = 1) \tag{4.87}$$

$$S_e = S_T - S_\beta \qquad (f = 5) \tag{4.88}$$

$$V_e = \frac{S_e}{5} \tag{4.89}$$

Using these results, we obtain the SN ratio and sensitivity:

$$n = 10 \log \frac{(1/r)\ (S_\beta - V_e)}{V_e} \tag{4.90}$$

$$S = 10 \log \frac{1}{r}\ (S_\beta - V_e) \tag{4.91}$$

Since we analyze based on a root square, the true value of S is equal to the estimation of β^2, that is, consumption of electricity needed for cutting, which should be smaller. If we perform smooth machining with a small amount of electricity, we can obviously improve dimensional accuracy and flatness.

2. *Improvement of work efficiency*. This analysis is not for machining performance but for work efficiency. In this case, an ideal function is expressed as

$$y = \beta T \tag{4.92}$$

The SN ratio and sensitivity are calculated as

$$S_T = y_{11}^2 + y_{12}^2 + \cdots + y_{23}^2 \qquad (f = 6) \tag{4.93}$$

$$S_\beta = \frac{(L_1 + L_2)^2}{2r} \qquad (f = 1) \tag{4.94}$$

$$S_{N\beta} = \frac{(L_1 - L_2)^2}{2r} \qquad (f = 1) \tag{4.95}$$

Now

$$L_1 = T_1 y_{11} + T_2 y_{12} + T_3 y_{13} \tag{4.96}$$

$$L_2 = T_1 y_{21} + T_2 y_{22} + T_3 y_{23} \tag{4.97}$$

$$r = T_1^2 + T_2^2 + T_3^2 \tag{4.98}$$

$$\eta = 10 \log \frac{(1/2r)\ (S_\beta - V_e)}{V_N} \tag{4.99}$$

$$S = 10 \log \frac{1}{2r}\ (S_\beta - V_e) \tag{4.100}$$

Then

$$V_e = \tfrac{1}{4}(S_T - S_\beta - S_{N\beta}) \tag{4.101}$$

$$V_N = \tfrac{1}{5}(S_T - S_\beta) \tag{4.102}$$

In this case, β^2 of the ideal function represents electricity consumption per unit time, which should be larger. To improve the SN ratio using equation (4.99) means to design a procedure of using a large amount of electricity smoothly. Since we minimize electricity consumption by equations (4.90) and (4.91), we can improve not only productivity but also quality by taking advantage of both analyses.

The efficiency ratio of y to M, which has traditionally been used by engineers, can neither solve quality problems nor improve productivity. On the other hand, the hourly machining amount (machining measurement) used by managers in manufacturing cannot settle noise problems and product quality in machining.

To improve quality and productivity at the same time, after measuring the data shown in Table 4.10 we should determine an optimal condition based on two values, SN ratio and sensitivity. To lower sensitivity in equation (4.91) is to minimize electric power per unit removal amount. If we cut more using a slight quantity of electricity, we can cut smoothly, owing to improvement in the SN ratio in equation (4.90), as well as minimize energy loss. In other words, it means to cut sharply, thereby leading to good surface quality after cutting and improved machining efficiency per unit energy.

When On and Off Conditions Exist

Since quality engineering is a generic technology, a common problem occurs for the mechanical, chemical, and electronic engineering fields. We explain the functionality evaluation method for on and off conditions using an example of machining.

Recently, evaluation methods of machining have made significant progress, as shown in an experiment implemented in 1997 by Ishikawajima–Harima Heavy Industries (IHI) and sponsored by the National Space Development Agency of Japan (NASDA). The content of this experiment was released in the Quality Engineering Symposium in 1998. Instead of using transformality [setting product dimensions input through a numerically controlled (NC) machine as signal and corresponding dimensions after machined as measurement characteristic], they measured input and output utilizing energy and work by taking advantage of the concept of generic function.

In 1959, the following experiment was conducted at the Hamamatsu Plant of the National Railway (now known as JR). In this experiment, various types of cutting tools (JIS, SWC, and new SWC cutting tools) and cutting conditions were assigned to an L_{27} orthogonal array to measure the net cutting power needed for a certain amount of cut. The experiment, based on 27 different combinations, revealed that the maximum power needed was several times as large as the minimum. This finding implies that excessive power may cause a rough surface or variability in dimensions. In contrast, cutting with quite a small amount of power means that we can cut sharply. Consequently, when we can cut smoothly and power effectively, material surfaces can be machined flat and variability in dimension can be reduced. In the Quality Engineering Symposium in June 1998, IHI released research on improvement in machining accuracy of a space rocket engine, entitled "Optimization of T_i High-Speed Machining Process" by Kenji Fujmoto and two others from Ishikawajima–Harima Heavy Industries, Kazuyuki Higashihara from the National Space Development Agency of Japan, and Kazuhito Takahashi from the University of Electro-Communications.

This research did not adopt transformality, which defines an ideal function as the following equation by setting the input signal of the NC machine N to signal M, and a dimension of a machined test piece y to output:

$$y = \beta M \tag{4.103}$$

The paradigm of the testing method in quality engineering maintains that we should select energy or work as input signal M and output y. The ideal function in equation (4.103) is not expressed by an equation proportional to energy, as IHI clearly enunciated in the forum. Since cutting itself is work in the cutting process, the amount of cut is regarded as work. Therefore, output is work. To make something work, we need energy or power. IHI originally wished to input electricity consumption. However, since it changes every second, they considered it inappropriate for use as signal. They conducted an experiment based on Table 4.11, in which T, M, and y indicate the number of tool paths, measured amount of cut, and electricity consumption, respectively. In addition, as noise factor levels, two different types of thickness of a flat plate N_1 and N_2, each of which also had different Young's modulus, were chosen.

The reason that they chose the amount of cut as signal M (as did the JR experiment in 1959) is that the amount of cut is regarded as required work. The ideal function was defined as follows, after they took the square root of both sides of the original equation:

$$\sqrt{y} = \beta \sqrt{M} \tag{4.104}$$

The only difference between (4.103) and (4.104) is that the former is not a generic but an objective function, whereas the latter is a generic function representing the relationship between work and energy. Since y represents a unit of energy, a square of y does not make sense from the viewpoint of physics. Thus, by taking the square root of both sides beforehand, we wish to make both sides equivalent to work and energy when they are squared.

On the other hand, unless the square of a measurement characteristic equals energy or work, we do not have any rationale in calculating the SN ratio based on a generic function. After taking the square root of both sides in equation (4.104),

Table 4.11
Cutting experiment

		T_1	T_2	T_3
N_1	M: amount of cut	M_{11}	M_{12}	M_{13}
	y: electricity consumption	y_{11}	y_{12}	y_{13}
N_2	M: amount of cut	M_{21}	M_{22}	M_{23}
	y: electricity consumption	y_{21}	y_{22}	y_{23}

we compute the total variation of output S_T, which is a sum of squares of the square root of each cumulative electricity consumption:

$$S_T = (\sqrt{y_{11}})^2 + (\sqrt{y_{12}})^2 + \cdots + (\sqrt{y_{23}})^2 = y_{11} + y_{12} + \cdots + y_{23} \qquad (f = 6) \tag{4.105}$$

This is equivalent to a total sum of electricity consumptions. Subsequently, total consumed energy S_T can be decomposed as follows:

$$\begin{aligned} S_T = {} & (\text{Electricity consumption used for cutting}) \\ & + (\text{electricity consumption used for loss or variability in cutting}) \end{aligned} \tag{4.106}$$

While an ideal relationship is defined by energy or work, we use the square root of both sides of the relationship in analyzing the SN ratio. After we take the square root of both sides, the new relationship is called a *generic function*.

We can compute the power used for cutting as a variation of proportional terms:

$$S_\beta = \frac{L_1 + L_2}{r_1 + r_2} \tag{4.107}$$

Now L indicates a linear equation using square roots and is calculated as

$$\begin{aligned} L_1 &= \sqrt{M_{11}}\sqrt{y_{11}} + \sqrt{M_{12}}\sqrt{y_{12}} + \sqrt{M_{13}}\sqrt{y_{13}} \\ L_2 &= \sqrt{M_{21}}\sqrt{y_{21}} + \sqrt{M_{22}}\sqrt{y_{22}} + \sqrt{M_{23}}\sqrt{y_{23}} \end{aligned} \tag{4.108}$$

$$\begin{aligned} r_1 &= (\sqrt{M_{11}})^2 + (\sqrt{M_{12}})^2 + (\sqrt{M_{13}})^2 \\ &= M_{11} + M_{12} + M_{13} \\ r_2 &= M_{21} + M_{22} + M_{23} \end{aligned} \tag{4.109}$$

Variation of sensitivity fluctuations $N\beta$,

$$S_{N\beta} \frac{L_1 - L_2}{r_1 + r_2} \tag{4.110}$$

According to Table 4.11, a general performance evaluation, or test of work quantity and quality, is simple to complete. All that we need to do in designing a real product is tuning, regarded as a relatively simple procedure. For any machine function, after defining an ideal function of signal and output based on work and energy, we should calculate the SN ratio and sensitivity by taking the square root of both sides of the functional equation.

4.4. Chemical Engineering

As a function regarding chemical reaction or molecules, often we can measure only percentage data of molecular conditions. Although we conduct R&D based on experiments on as small a scale as possible (e.g., flasks), the ultimate goal is optimal design of large-scale production processes. Quality engineering in chemical reactions deals with a technical method of studying functional robustness on a small scale. Robustness of chemical reactions revolves around control of reaction speed.

Function of an Engine

The genetic function of an engine is a chemical reaction. Our interest in oxygen aspirated into an engine is in how it is distributed in the exhaust.

- *Insufficient reaction.* We set the fraction of oxygen contained in CO_2 and CO to p.
- *Sufficient reaction.* We set the fraction of oxygen contained in CO_2 to q.
- *Side reaction* (e.g., of N_{ox}). We set the fraction of oxygen contained in side-reacted substances to $1 - p - q$.

Based on this definition, ideally the chemical reaction conforms to the following exponential equations:

$$p = e^{-\beta_1 T} \tag{4.111}$$

$$p + q = e^{-\beta_2 T} \tag{4.112}$$

Here T represents the time needed for one cycle of the engine and p and q are measurements in the exhaust. The total reaction rate β_1 and side reaction rate β_2 are calculated as

$$\beta_1 = \frac{1}{T} \ln \frac{1}{p} \tag{4.113}$$

$$\beta_2 = \frac{1}{T} \ln \frac{1}{p + q} \tag{4.114}$$

It is desirable that β_1 become greater and β_2 be close to zero. Therefore, we compute the SN ratio as

$$\eta = 10 \log \frac{\beta_1^2}{\beta_2^3} \tag{4.115}$$

The sensitivity is

$$S = 10 \log \beta_1^2 \tag{4.116}$$

The larger the sensitivity, the more the engine's output power increases, the larger the SN ratio becomes, and the less effect the side reaction has.

Cycle time T should be tuned such that benefits achieved by the magnitude of total reaction rate and loss by the magnitude of side reaction are balanced optimally.

Selecting noise N (which has two levels, such as one for the starting point of an engine and the other for 10 minutes after starting) as follows, we obtain the data shown in Table 4.12. According to the table, we calculate the reaction rate for N_1 and N_2. For an insufficient reaction,

$$\beta_{11} = \frac{1}{T} \ln \frac{1}{p_1} \tag{4.117}$$

$$\beta_{12} = \frac{1}{T} \ln \frac{1}{p_2} \tag{4.118}$$

Table 4.12
Function of engine based on chemical reaction

	N_1	N_2
Insufficient reaction p	p_1	p_2
Objective reaction q	q_1	q_2
Total	$p_1 + q_1$	$p_2 + q_2$

For a side reaction,

$$\beta_{21} = \frac{1}{T} \ln \frac{1}{p_1 + q_1} \tag{4.119}$$

$$\beta_{22} = \frac{1}{T} \ln \frac{1}{p_2 + q_2} \tag{4.120}$$

Since the total reaction rates β_{11} and β_{12} are larger-the-better, their SN ratios are as follows:

$$\eta_1 = -10 \log \frac{1}{2} \left(\frac{1}{\beta_{11}^2} + \frac{1}{\beta_{12}^2} \right) \tag{4.121}$$

Since the side reaction rates β_{21} and β_{22} are smaller-the-better, their SN ratio is

$$\eta_2 = -10 \log \tfrac{1}{2}(\beta_{21}^2 + \beta_{22}^2) \tag{4.122}$$

Finally, we compute the total SN ratio and sensitivity as follows:

$$\eta = \eta_1 + \eta_2 \tag{4.123}$$

$$S = \eta_1 \tag{4.124}$$

General Chemical Reactions

If side reactions barely occur and we cannot trust their measurements in a chemical reaction experiment, we can separate a point of time for measuring the total reaction rate $(1 - p)$ and a point of time for measuring the side reaction rate $(1 - p - q)$. For example, we can set T_1 to 1 minute and T_2 to 30 minutes, as illustrated in Table 4.13. When $T_1 = T_2$, we can use the procedure described in the preceding section.

Table 4.13
Experimental data

	T_1	T_2
Insufficient reaction p	p_{11}	p_{12}
Objective reaction q	q_{11}	q_{12}
Total	$p_{11} + q_{11}$	$p_{12} + q_{12}$

If the table shows that $1 - p_{11}$ is more or less than 50% and $1 - (p_{12} + q_{12})$ ranges at least from 10 to 50%, this experiment is regarded as good enough. p_{11} and $(p_{12} + q_{12})$ are used for calculating total reaction and side reaction rates, respectively:

$$\beta_1 = \frac{1}{T} \ln \frac{1}{p_{11}} \tag{4.125}$$

$$\beta_2 = \frac{1}{T_2} \ln \frac{1}{p_{12} + q_{12}} \tag{4.126}$$

Using the equations above, we can compute the SN ratio and sensitivity as

$$\eta = 10 \log \frac{\beta_1^2}{\beta_1^2} \tag{4.127}$$

$$S = 10 \log \beta_1^2 \tag{4.128}$$

Evaluation of Images

An *image* represents a picture such as a landscape transformed precisely on each pixel of a flat surface. Sometimes an image of a human face is whitened compared to the actual one; however, quality engineering does not deal with this type of case because it is an issue of product planning and tuning.

In a conventional research method, we have often studied a pattern of three primary colors (including a case of decomposing a gray color into three primary colors, and mixing up three primary colors into a gray color) as a test pattern. When we make an image using a pattern of three primary colors, a density curve (a common logarithm of a reciprocal of permeability or reflection coefficient) varies in accordance with various conditions (control factors). By taking measurements from this type of density curve, we have studied an image. Although creating a density curve is regarded as reasonable, we have also measured D_{max}, D_{min}, and gamma from the curve. In addition to them (e.g., in television), resolution and image distortion have been used as measurements.

Because the consumers' demand is to cover a minimum density difference of three primary colors (according to a filmmaker, this is the density range 0 to 10,000) that can be recognized by a human eye's photoreceptor cell, the resolving power should cover up to the size of the light-sensitive cell. However, quality engineering does not pursue such technical limitations but focuses on improving imaging technology. That is, its objective is to offer evaluation methods to improve both quality and productivity.

For example, quality engineering recommends the following procedure for taking measurements:

1. *Condition* M_1. Create an image of a test pattern using luminosity 10 times as high and exposure time one-tenth as high as their current levels.
2. *Condition* M_2. Create an image of a test pattern using as much luminosity and exposure time as their current levels.
3. *Condition* M_3. Create an image of a test pattern using luminosity one-tenth as high and exposure time 10 times as high as their current levels.

At the three sensitivity curves, we select luminosities corresponding to a permeability or reflection coefficient of 0.5. For density, it is 0.301. For a more practical experiment, we sometimes select seven levels for luminosity, such as 1000, 100, 10, 1, 1/10, 1/100, and 1/1000 times as much as current levels. At the same time, we choose exposure time inversely proportional to each of them.

After selecting a logarithm of exposure time E for the value 0.301, the SN ratio and sensitivity are calculated for analysis.

Next, we set a reading of exposure time E (multiplied by a certain decimal value) for each of the following seven levels of signal M (logarithm of luminosity) to $y_1, y_2, \ldots, y_7$:

$$M_1 = -3.0, \quad M_2 = -2.0, \quad \ldots, \quad M_7 = +3.0 \tag{4.129}$$

This is a procedure for calculating a necessary image-center luminosity from a sensitivity curve by taking into account a range of consumer uses. This procedure is considered appropriate from the standpoint not only of quality engineering but also the performance evaluation of photo film.

The ASA valve of a photo film indicates a necessary luminosity at the center of an image for each film. For a permeability or reflection coefficient, a point with value 0.5 is easiest to measure and the most appropriate for evaluation of an image, whereas we do not create images at D_{max} and D_{min}. Quality engineering is a strategic technology used to evaluate performance by taking consumer uses into account. Consumer uses are all signals and noises. In the case of an image, a signal is color density. To study an image by breaking down each of the miscellaneous color densities into a pattern of three primary colors is a rationalization of the measurement method.

The reason for measuring data at the center (at the point where the reflection or permeability coefficient is 50%) is that we wish to avoid inaccurate measurement, and this method has long been used in the telecommunication field. In addition, the reason for taking luminosity and exposure time in inverse proportion to this to see the density range is that we wish to check the linearity of color density of an image for luminosity.

This linearity is totally different from the study of images by pixels. That is, in dealing with digital information, we can select only the accuracy of each dot as a measurement because total image performance improves as the accuracy of each dot increases.

Among consumer uses are not only signals but also noises, which should be smaller to enhance performance. For a photographic film, conditions regarding film preservation and environment when a picture is taken are included in noises. We select the following levels of film preservation conditions as two noise levels:

N_1: good preservation condition, where the density change is small

N_2: poor preservation condition, where the density change is large

In this case, a small difference between the two sensitivity curves for N_1 and N_2 represents a better performance. A good way to designing an experiment is to compound all noise levels into only two levels. Thus, no matter how many noise levels we have, we should compound them into two levels.

Since we have three levels, K_1, K_2, and K_3, for a signal of three primary colors, two levels for a noise, and seven levels for an output signal, the total number of data is $(7)(3)(2) = 42$. For each of the three primary colors, we calculate the SN ratio and sensitivity. Now we show only the calculation of K_1. Based on Table 4.14,

Table 4.14
Experimental data

		M_1	M_2	···	M_7	Linear Equation
K_1	N_1	y_{11}	y_{12}	···	y_{17}	L_1
	N_2	y_{21}	y_{22}	···	y_{27}	L_2
K_2	N_1	y_{31}	y_{32}	···	y_{37}	L_3
	N_2	y_{41}	y_{42}	···	y_{47}	L_4
K_3	N_1	y_{51}	y_{52}	···	y_{57}	L_5
	N_2	y_{61}	y_{62}	···	y_{67}	L_6

we proceed with the calculation. Now, by taking into account that M has seven levels, we view $M_4 = 0$ as a standard point and subtract the value of M_4 from M_1, M_2, M_3, M_5, M_6, and M_7, each of which is set to y_1, y_2, y_3, y_5, y_6, and y_7.

By subtracting a reading y for $M_4 = 0$ in case of K_1, we obtain the following linear equations:

$$L_1 = -3y_{11} - 2y_{12} - \cdots + 3y_{17} \tag{4.130}$$

$$L_2 = -\ 3y_{21} - \cdots + 3y_{27} \tag{4.131}$$

Based on these, we define $r = (-3)^2 + (-2)^2 + \cdots + 3^2 = 28$.

$$S_\beta = \frac{(L_1 + L_2)^2}{2r} \tag{4.132}$$

$$S_{N\beta} = \frac{(L_1 - L_2)^2}{2r} \tag{4.133}$$

$$S_T = y_{11}^2 + y_{12}^2 + \cdots + y_{27}^2 \tag{4.134}$$

$$V_e = \tfrac{1}{12}(S_T - S_\beta - S_{N\beta}) \tag{4.135}$$

$$V_N = \tfrac{1}{13}(S_T - S_\beta) \tag{4.136}$$

$$\eta_1 = 10 \log \frac{(1/2r)\ (S_\beta - V_e)}{V_N} \tag{4.137}$$

$$S_1 = 10 \log \frac{1}{2r} (S_\beta - V_e) \tag{4.138}$$

As for K_2 and K_3, we calculate η_2, S_2, η_3, and S_3. The total SN ratio η is computed as the sum of η_1, η_2, and η_3. To balance densities ranging from low to high for K_1, K_2, and K_3, we should equalize sensitivities for three primary colors, S_1, S_2, and S_3, by solving simultaneous equations based on two control factors. Solving them such that

$$S_1 = S_2 = S_3 \tag{4.139}$$

holds true under standard conditions is not a new method.

The procedure discussed thus far is the latest method in quality engineering, whereas we have used a method of calculating the density for a logarithm of luminosity log *E*. The weakness of the latter method is that a density range as output is different at each experiment. That is, we should not conduct an experiment using the same luminosity range for films of ASA 100 and 200. For the latter, we should perform an experiment using half of the luminosity needed for the former. We have shown the procedure to do so.

Quality engineering is a method of evaluating functionality with a single index η. Selection of the means for improvement is made by specialized technologies and specialists. Top management has the authority for investment and personnel affairs. The result of their performance is evaluated based on a balance sheet that cannot be manipulated. Similarly, although hardware and software can be designed by specialists, those specialists cannot evaluate product performance at their own discretion.

Functionality Such as Granulation or Polymerization Distribution

If you wish to limit granulation distribution within a certain range, you must classify granules below the lower limit as excessively granulated, ones around the center as targeted, and ones above the upper limit as insufficiently granulated, and proceed with analysis by following the procedure discussed earlier for chemical reactions. For polymerization distribution, you can use the same technique.

Separation System

Consider a process of extracting metallic copper from copper sulfide included in ore. When various types of substances are mixed with ore and the temperature of a furnace is raised, deoxidized copper melts, flows out of the furnace, and remains in a mold. This is called *crude copper ingot*. Since we expect to extract 100% of the copper contained in ore and convert it into crude copper (product), this ratio of extraction is termed *yield*. The percentage of copper remaining in the furnace slag is regarded as the loss ratio p. If $p = 0$, the yield becomes 100%. Because in this case we wish to observe the ratio at a single point of time during reaction, calibration will be complicated.

On the other hand, as the term *crude copper* implies, there exist a considerable number (approximately 1 to 2% in most cases) of ingredients other than copper. Therefore, we wish to bring the ratio of impurity contained in copper ingot, q^*, close to zero. Now, setting the mass of crude copper to A (kilograms), the mass of slag to B (kilograms) (this value may not be very accurate), the ratio of impurity in crude copper to q^*, and the ratio of copper in slag to p^*, we obtain Table 4.15 for input/output. From here we calculate the following two error ratios, p and q:

$$p = \frac{Bp^*}{A(1 - q^*) + Bp^*} \tag{4.140}$$

$$q = \frac{Aq^*}{Aq^* + B(1 - p^*)} \tag{4.141}$$

The error ratio p represents the ratio of copper molecules that is originally contained in ore but mistakenly left in the slag after smelting. Then $1 - p$ is called the *yield*. Subtracting this yield from 1, we obtain the error ratio p. The error ratio q indicates the ratio of all noncopper molecules that is originally contained in the furnace but mistakenly included in the product or crude copper. Both ratios are calculated as a mass ratio. Even if the yield $1 - p$ is large enough, if the error

Table 4.15
Input/output for copper smelting

Input	Output: Product	Slag	Total
Copper	$A(1 - q^*)$	Bp^*	$A(1 - q^*) + Bp^*$
Noncopper	Aq^*	$B(1 - p^*)$	$Aq^* + B(1 - p^*)$
Total	A	B	$A + B$

ratio q is also large, this smelting is considered inappropriate. After computing the two error ratios p and q, we prepare Table 4.16.

Consider the two error ratios p and q in Table 4.15. If copper is supposed to melt well and move easily in the product when the temperature in the furnace is increased, the error ratio p decreases. However, since ingredients other than copper melt well at the same time, the error ratio q rises. A factor that can decrease p and increase q is regarded as an adequate variable for tuning. Although most factors have such characteristics, more or less, we consider it real technology to reduce errors regarding both p and q rather than obtaining effects by tuning. In short, this is smelting technology with a large functionality.

To find a factor level reducing both p and q for a variety of factors, after making an adjustment so as not to change the ratio of p to q, we should evaluate functionality. This ratio is called the *standard SN ratio*. Since gain for the SN ratio accords with that when $p = q$, no matter how great the ratio of p to q, in most cases we calculate the SN ratio after p is adjusted equal to q. Primarily, we compute p_0 as follows when we set $p = q = p_0$:

$$p_0 = \frac{1}{1 + \sqrt{[(1/p) - 1][(1/q) - 1]}} \tag{4.142}$$

Secondarily, we calculate the standard SN ratio η as

$$\eta = 10 \log \frac{(1 - 2p_0)^2}{4p_0(1 - p_0)} \tag{4.143}$$

The details are given in Reference 5.

Once the standard SN ratio η is computed, we determine an optimal condition according to the average SN ratios for each control factor level, estimate SN ratios

Table 4.16
Input/output expressed by error ratios

Input	Output: Product	Slag	Total
Copper	$1 - p$	p	1
Noncopper	q	$1 - q$	1
Total	$1 - p + q$	$1 + p - q$	2

for optimal and initial conditions, and calculate gains. After this, we conduct a confirmatory experiment for the two conditions, compute SN ratios, and compare estimated gains with those obtained from this experiment.

Based on the experiment under optimal conditions, we calculate p and q. Unless a sum of losses for p and q is minimized, timing is set using factors (such as the temperature in the furnace) that influence sensitivity but do not affect the SN ratio. In tuning, we should gradually change the adjusting factor level. This adjustment should be made after the optimal SN ratio condition is determined.

The procedure detailed here is even applicable to the removal of harmful elements or the segregation of garbage.

4.5. Medical Treatment and Efficacy Experimentation

A key objective of quality engineering is to rationalize functional evaluation. For example, for medical efficacy, it proposes a good method of evaluating great effect and small side effects in a laboratory. Evaluation in a laboratory clarifies whether we can assess both main effects and side effects without clinical tests on patients at the same time. We discuss monitoring of individual patients later in this section.

❒ Example

First, we consider a case of evaluating the main effects on a cancer cell and the side effects on a normal cell. Although our example is an anticancer drug, this analytic procedure holds true for thermotherapy and radiation therapy using supersonic or electromagnetic waves. If possible, by taking advantage of an L_{18} orthogonal array, we should study the method using a drug and such therapies at the same time.

Quality engineering recommends that we experiment on cells or animals. We need to alter the density of a drug to be assessed by h milligrams (e.g., 1 mg) per unit time (e.g, 1 minute). Next, we select one cancer cell and one normal cell and designate them M_1 and M_2, respectively. In quality engineering, M is called a *signal factor*. Suppose that M_1 and M_2 are placed in a certain solution (e.g., water or a salt solution), the density is increased by 1 mg/minute, and the length of time that each cell survives is measured. Imagine that the cancer and normal cells die at the eighth and fourteenth minutes, respectively. We express these data as shown in Table 4.17, where 1 indicates "alive" and 0 indicates "dead." In addition, M_1 and

Table 4.17
Data for a single cell

	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}	T_{13}	T_{14}	T_{15}
M_1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
M_2	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0

M_2 are cancer and normal cells, and T_1, T_2, ... indicate each lapse of time: 1, 2, ... minutes.

For digital data regarding a dead-or-alive (or cured-or-not cured) state, we calculate LD_{50} (lethal dose 50). In this case, the LD_{50} value of M_1 is 7.5 because a cell dies between T_7 and T_8, whereas that of M_2 is 13.5. LD_{50} represents the quantity of drug needed until 50% of cells die.

In quality engineering, both M and T are signal factors. Although both signal and noise factors are variables in use, a signal factor is a factor whose effect should exist, and conversely, a noise is a factor whose effect should be minimized. That is, if there is no difference between M_1 and M_2, and furthermore, between each different amount of drug (or of time in this case), this experiment fails. Then we regard both M and T as signal factors. Thus, we set up X_1 and X_2 as follows:

$$X_1 = LD_{50} \text{ of a cancer cell} \tag{4.144}$$

$$X_2 = LD_{50} \text{ of a normal cell} \tag{4.145}$$

In this case, the smaller X_1 becomes, the better the drug's performance becomes. In contrast, X_2 should be greater. Quality engineering terms the former the *smaller-the-better characteristic* and the latter *larger-the-better characteristic*. The following equation calculates the SN ratio:

$$\eta = 20 \log \frac{X_2}{X_1} \tag{4.146}$$

In this case, since there is no noise, to increase η is equivalent to enlarging the ratio X_2/X_1.

For a more practical use, we often test N_1, N_2, ... , N_k as k cancer cells, and N'_1, N'_2, ... , N'_k as k′ normal cells. Suppose that $k = 3$ and $k' = 3$. N_1, N_2, and N_3 can be selected as three different types of cancer cells or three cancer cells of the same type. Or we can choose two cancer cells of the same type and one cancer cell of a different type.

Next, we obtain dead-or-alive data, as illustrated in Table 4.18, after placing three cells for each of M_1 and M_2. Table 4.19 shows the LD_{50} values of N_1, N_2, and N_3 and N'_1, N'_2, and N'_3.

Table 4.18
Dead-or-alive data

		T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}	T_{13}	T_{14}	T_{15}	T_{16}	T_{17}	T_{18}	T_{19}	T_{20}
M_1	N_1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	N_2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	N_3	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
M_2	$N_{1'}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
	$N_{2'}$	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
	$N_{3'}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0

Table 4.19
LD_{50} data

M_1	5.5	3.5	11.5
M_2	14.5	8.5	19.5

Since the data for M_1 should have a smaller standard deviation as well as a smaller average, after calculating an average of a sum of squared data, we multiply its logarithm by 10. This is termed the *smaller-the-better SN ratio:*

$$\eta_1 = -10 \log \tfrac{1}{3}(5.5^2 + 3.5^2 + 11.5^2) = -17.65 \text{ dB} \tag{4.147}$$

On the other hand, because the LD_{50} value of a normal cell M_2 should be infinitesimal, after computing an average sum of reciprocal data, we multiply its logarithm by 10. This is *larger-the-better SN ratio*:

$$\eta_2 = -10 \log \frac{1}{3}\left(\frac{1}{14.5^2} + \frac{1}{8.5^2} + \frac{1}{19.5^2}\right) = 21.50 \text{ dB} \tag{4.148}$$

Therefore, by summing up η_1 and η_2, we calculate the following η, which is the total SN ratio integrating both main and side effects:

$$\begin{aligned}\eta &= \eta_1 + \eta_2 \\ &= -17.65 + 21.50 \\ &= 3.85 \text{ dB}\end{aligned} \tag{4.149}$$

For example, given two drugs A_1 and A_2, we obtain the experimental data shown in Table 4.20. If we compare the drugs, N_1, N_2, and N_3 for A_1 should be consistent with N_1', N'_2, and N_3' for A_2. Now suppose that the data of A_1 are the same as those in Table 4.19 and that a different drug A_2 has LD_{50} data for M_1 and M_2 as illustrated in Table 4.20. A_1's SN ratio is as in equation (4.149). To compute A_2's SN ratio, we calculate the SN ratio for the main effect η_1 as follows:

$$\eta_1 = -10 \log \tfrac{1}{3}(18.5^2 + 11.5^2 + 20.5^2) = -24.75 \text{ dB} \tag{4.150}$$

Table 4.20
Data for comparison experiment

A_1	M_1	5.5	3.5	11.5
	M_2	14.5	8.5	19.5
A_2	M_1	18.5	11.5	20.5
	M_2	89.5	40.5	103.5

Next, as a larger-the-better SN ratio, we calculate the SN ratio for the side effects η_2 as

$$\eta_2 = -10 \log \frac{1}{3}\left(\frac{1}{89.5^2} + \frac{1}{40.5^2} + \frac{1}{103.5^2}\right) = 35.59 \qquad (4.151)$$

Therefore, we obtain the total SN ratio, combining the main and side effects, by adding them up as follows:

$$\eta = -24.75 + 35.59 = 10.84 \qquad (4.152)$$

Finally, we tabulate these results regarding comparison of A_1 and A_2 in Table 4.21, which is called a *benchmarking test*. What we find out from Table 4.21 is that the side effect of A_2 is larger than A_1's by 14.09 dB, or 25.6 times, whereas the main effect of A_2 is smaller than A_1's by 7.10 dB, or 1/5.1 times. On balance, A_2's effect is larger than A_1's by 6.99 dB, or 5 times. This fact reveals that if we increase an amount of A_2, we can improve both main and side effects by 3.49 dB compared to A_1's. That is, the main effect is enhanced 2.3 times, and the side effect is reduced 1/2.3 times. By checking the SN ratio using the operating window method, we can know whether the main and side-effects are improved at the same time.

When we study therapy such as using heat or electromagnetic waves, we select as many control factors as possible, assign them to an orthogonal array L_{18}, increase the power at a certain interval, and measure data as shown in Table 4.18. Then, after preparing a table such as Table 4.19 for each experiment in the orthogonal array, we calculate the SN ratios. As for remaining calculations, we can follow the normal quality engineering procedures.

When we wish to measure temperature itself when dealing with thermotherapy, we should choose a temperature difference from a standard body temperature. For example, suppose that we set the starting temperature to a body temperature (e.g., 36°C) when we raise the temperature of a Schale or petri dish containing one cancer cell M_1 and one normal cell M_2 by 1° each 10 minutes (or 5 minutes). Indeed, it is possible to raise the temperature quickly; however, by increasing it slowly, we should monitor how many times we need to raise the temperature until a cell dies. In this case, the temperature T is regarded as a signal whose standard

Table 4.21
Comparison of drugs A_1 and A_2

	Main Effect	Side Effect	Total
A_1	−17.65	21.50	3.85
A_2	−24.75	35.59	10.84
Gain	−7.10	14.09	6.99

point is 36°C. If the cancer cell dies at 43°C, the LD_{50} value is 6.5°C, the difference from the standard point of 36°C. If the normal cell dies at 47°C, the LD_{50} value is 10.5°C. When a sonic or electromagnetic wave is used in place of temperature, we need to select the wave's power (e.g., raise the power by 1 W/minute.

4.6. Software Testing

Quality engineering classifies a signal type into two categories, the quality of an active signal (in the case of software that a user uses actively) and the quality of the MTS (Mahalanobis–Taguchi system), dealing with a passive signal. In this section we detail the former, and in the next section we discuss the case for the latter.

Two Types of Signal Factor and Software

Quality engineering supposes that all conditions of use necessarily belong to a signal or noise factor. Apart from a condition where a user uses it actively or passively, the effects of a signal factor are those that are essential. When we plan to design software (software product) using a system with computers, the software is considered a user's *active signal factor*. On the other hand, when we conduct inspection, diagnosis, or prediction using research data (including various sensing data), the entire group of data is regarded as consisting of *passive signal factors*. Signal factors should not only have a common function under various conditions of use but also have small errors.

In this section we explain how to use the SN ratio in taking measures against bugs when we conduct a functional test on software that a user uses actively. Software products have a number of active signal factors and consist of various levels of signal factors. In this section we discuss ways to measure and analyze data to find bugs in software.

Layout of Signal Factors in an Orthogonal Array

In testing software, there is a multiple-step process, that is, a number of signals. In quality engineering, we do not critque software design per se, but discuss measures that enable us to find bugs in designed software. We propose a procedure for checking whether software contains bugs or looking for problems with (diagnosis of) bugs.

As we discussed before, the number of software signal factors is equivalent to the number of steps involved. In actuality, the number of signal factors is tremendous, and the number is completely different in each factor level. With software, how large an orthogonal array we can use should be discussed even when signal factors need to be tested at every step or data can be measured at a certain step. To use an L_{36} orthogonal array repeatedly is one of the practical methods we can use, as shown next.

- ❏ *Procedure 1.* We set the number of multilevel signal factors to k. If k is large, we select up to 11 signal factors of two levels and up to 12 factors of three levels. Therefore, if $k \leq 23$, we should use an L_{36} orthogonal array. If k is more than 23, we should use a greater orthogonal array (e.g., if $24 \leq k \leq 59$, L_{108} is to be used) or use an L_{36} orthogonal array repeatedly to allocate all factors. We should lay out all signal factors as an array after reducing the number of each factor level to two or three.

- *Procedure 2.* We conduct a test at 36 different conditions that are laid out on an orthogonal array. If we obtain an acceptable result for each experiment in the orthogonal array, we record a 0; conversely, if we do not obtain an acceptable result, we record a 1. Basically by looking at output at the final step, we judge by 0 or 1. Once all data (all experiments in case of L_{36}) are zero, the test is completed. If even one datum of 1 remains in all experimental combination, the software is considered to have bugs.
- *Procedure 3.* As long as there are bugs, we need to improve the software design. To find out at what step a problem occurs when there are a number of bugs, we can measure intermediate data. In cases where there are a small number of bugs, we analyze interactions. Its application to interactions is discussed in Section 4.7.

Software Diagnosis Using Interaction

For the sake of convenience, we explain the procedure by using an L_{36} orthogonal array; this also holds true for other types of arrays. According to procedure 3 in the preceding section, we allocate 11 two-level factors and 12 three-level factors to an L_{36} orthogonal array, as shown in Table 4.22. Although one-level factors are not assigned to the orthogonal array, they must be tested.

We set two-level signal factors to A, B, ... , and K, and three-level signal factors to L, M, ... , and W. Using the combination of experiments illustrated in Table 4.23, we conduct a test and measure data about whether or not software functions properly for all 36 combinations. When we measure output data for tickets and changes in selling tickets, we are obviously supposed to set 0 if both outputs are correct and 1 otherwise. In addition, we need to calculate an interaction for each case of 0 and 1.

Now suppose that the measurements taken follow Table 4.23. In analyzing data, we sum up data for all combinations of A to W. The total number of combinations is 253, starting with AB and ending with VW. Since we do not have enough space to show all combinations, for only six combinations of AB, AC, AL, AM, LM, and LW, we show a two-way table, Table 4.23. For practical use, we can use a computer to create a two-way table for all combinations.

Table 4.23 includes six two-way tables for *AB, AC, AL, AM, LM, and LW*, and each table comprises numbers representing a sum of data 0 or 1 for four conditions of A_1B_1 (Nos. 1–9), A_1B_2 (Nos. 10–18), A_2B_1 (Nos. 19–27), and A_2B_2 (Nos. 28–36). Despite all 253 combinations of two-way tables, we illustrate only six. After we create a two-dimensional table for all possible combinations, we need to consider the results.

Based on Table 4.23, we calculate combined effects. Now we need to create two-way tables for combinations whose error ratio is 100% because if an error in software exists, it leads to a 100% error. In this case, using a computer, we should produce tables only for L_2 and L_3 for A_2, M_3 for A_2, M_2 for L_3, W_2 and W_3 for L_2, and W_2 and W_3 for L_3. Of 253 combinations of two-dimensional tables, we output only 100% error tables, so we do not need to output those for AB and AC.

We need to enumerate not only two-way tables but also all 100% error tables from all possible 253 tables because bugs in software are caused primarily by combinations of signal factor effects. It is software designers' job to correct errors.

Table 4.22
Layout and data of L_{36} orthogonal array

No.	*A* 1	*B* 2	*C* 3	*D* 4	*E* 5	*F* 6	*G* 7	*H* 8	*I* 9	*J* 10	*K* 11	*L* 12	*M* 13	*N* 14	*O* 15	*P* 16	*Q* 17	*R* 18	*S* 19	*T* 20	*U* 21	*V* 22	*W* 23	Data
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
2	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	1
3	1	1	1	1	1	1	1	1	1	1	1	3	3	3	3	3	3	3	3	3	3	3	3	1
4	1	1	1	1	1	2	2	2	2	2	2	1	1	1	1	2	2	2	2	3	3	3	3	0
5	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	3	3	3	3	1	1	1	1	0
6	1	1	1	1	1	2	2	2	2	2	2	3	3	3	3	1	1	1	1	2	2	2	2	1
7	1	1	2	2	2	1	1	1	2	2	2	1	1	2	3	1	2	3	3	1	2	2	3	0
8	1	1	2	2	2	1	1	1	2	2	2	2	2	3	1	2	3	1	1	2	3	3	1	0
9	1	1	2	2	2	1	1	1	2	2	2	3	3	1	2	3	1	2	2	3	1	1	2	1
10	1	2	1	2	2	1	2	2	1	1	2	1	1	3	2	1	3	2	3	2	1	3	2	0
11	1	2	1	2	2	1	2	2	1	1	2	2	2	1	3	2	1	3	1	3	2	1	3	1
12	1	2	1	2	2	1	2	2	1	1	2	3	3	2	1	3	2	1	2	1	3	2	1	0
13	1	2	2	1	2	2	1	2	1	2	1	1	2	3	1	3	2	1	3	3	2	1	2	0
14	1	2	2	1	2	2	1	2	1	2	1	2	3	1	2	1	3	2	1	1	3	2	3	1
15	1	2	2	1	2	2	1	2	1	2	1	3	1	2	3	2	1	3	2	2	1	3	1	0
16	1	2	2	2	1	2	2	1	2	1	1	1	2	3	2	1	1	3	2	3	3	2	1	0
17	1	2	2	2	1	2	2	1	2	1	1	2	3	1	3	2	2	1	3	1	1	3	2	1
18	1	2	2	2	1	2	2	1	2	1	1	3	1	2	1	3	3	2	1	2	2	1	3	1
19	2	1	2	2	1	1	2	2	1	2	1	1	2	1	3	3	3	1	2	2	1	2	3	0
20	2	1	2	2	1	1	2	2	1	2	1	2	3	2	1	1	1	2	3	3	2	3	1	1
21	2	1	2	2	1	1	2	2	1	2	1	3	1	3	2	2	2	3	1	1	3	1	2	1
22	2	1	2	1	2	2	2	1	1	1	2	1	2	2	3	3	1	2	1	1	3	3	2	0
23	2	1	2	1	2	2	2	1	1	1	2	2	3	3	1	1	2	3	2	2	1	1	3	1
24	2	1	2	1	2	2	2	1	1	1	2	3	1	1	2	2	3	1	3	3	2	2	1	1

Table 4.22 (*Continued*)

No.	*A* 1	*B* 2	*C* 3	*D* 4	*E* 5	*F* 6	*G* 7	*H* 8	*I* 9	*J* 10	*K* 11	*L* 12	*M* 13	*N* 14	*O* 15	*P* 16	*Q* 17	*R* 18	*S* 19	*T* 20	*U* 21	*V* 22	*W* 23	Data
25	2	1	1	2	2	2	1	2	2	1	1	1	3	2	1	2	3	3	1	3	1	2	2	0
26	2	1	1	2	2	2	1	2	2	1	1	2	1	3	2	3	1	1	2	1	2	3	3	1
27	2	1	1	2	2	2	1	2	2	1	1	3	2	1	3	1	2	2	3	2	3	1	1	1
28	2	2	2	1	1	1	1	2	2	1	2	2	1	3	3	3	2	2	1	3	1	2	1	1
29	2	2	2	1	1	1	1	2	2	1	2	2	1	3	3	3	2	2	1	3	1	2	1	1
30	2	2	2	1	1	1	1	2	2	1	2	3	2	1	1	1	3	3	2	1	2	3	2	1
31	2	2	1	2	1	2	1	1	1	2	2	1	3	3	3	2	3	2	2	1	2	1	1	0
32	2	2	1	2	1	2	1	1	1	2	2	2	1	1	1	3	1	3	3	2	3	2	2	1
33	2	2	1	2	1	2	1	1	1	2	2	3	2	2	2	1	2	1	1	3	1	3	3	1
34	2	2	1	1	2	1	2	1	2	2	1	1	3	1	2	3	2	3	1	2	2	3	1	1
35	2	2	1	1	2	1	2	1	2	2	1	2	1	2	3	1	3	1	2	3	3	1	2	1
36	2	2	1	1	2	1	2	1	2	2	1	3	2	3	1	2	1	2	3	1	1	2	3	1

Table 4.23
Supplemental tables

(1) *AB* two-way table	B_1	B_2	Total
A_1	4	4	8
A_2	6	8	14
Total	10	12	22

(2) *AC* two-way table	C_1	C_2	Total
A_1	4	4	8
A_2	7	7	14
Total	11	11	22

(3) *AL* two-way table	L_1	L_2	L_3	Total
A_1	0	4	4	8
A_2	2	6	6	14
Total	2	10	10	22

(4) *AM* two-way table	M_1	M_2	M_3	Total
A_1	1	2	5	8
A_2	6	4	4	14
Total	7	6	9	22

(5) *LM* two-way table	M_1	M_2	M_3	Total
L_1	0	0	2	2
L_2	4	2	4	10
L_3	3	4	3	10
Total	7	6	9	22

(6) *LW* two-way table	W_1	W_2	W_3	Total
L_1	1	0	1	2
L_2	2	4	4	10
L_3	2	4	4	10
Total	5	8	9	22

Once they modify 100% errors, they perform a test again based on an L_{36} orthogonal array. If 100% error combinations remain, they correct them again. Although this procedure sounds imperfect, it quite often streamlines the debugging task many times over.

System Decomposition

When, by following the method based on Table 4.22, we find it extremely difficult to seek root causes because of quite a few bugs, it is more effective for debugging to break signal factors into small groups instead of selecting all of them at a time. Some ways to do so are described below.

1. After splitting signal factors into two groups, we lay out factors in each of them to an L_{18} orthogonal array. Once we correct all bugs, we repeat a test based on an L_{36} orthogonal array containing all signal factors. We can reduce bugs drastically by two L_{18} tests, thereby simplifying the process of seeking root causes in an L_{36} array with few bugs.
2. When we are faced with difficulties finding causes in an L_{36} array because of too many bugs, we halve the number of signal factors by picking alternative factors, such as A, C, E, G, ... , W, and check these bugs. Rather than selecting every other factor, we can choose about half that we consider important. The fact that the number of bugs never changes reveals that all bugs are caused by a combination of about half the signal factors. Then we halve the number of factors. Until all root causes are detected, we continue to follow this process. In contrast, if no bug is found in the first-half combinations of signal factors, we test the second half. Further, if there happens to be no bug in this test, we can conclude that interactions between the first and second halves generate bugs. To clarify causes, after dividing each of the first and second halves into two groups, we investigate all four possible combinations of the two groups.
3. From Table 4.22, by correcting 100% errors regarding A, L, and M based on Table 4.23, we then check for bugs in an L_{36} orthogonal array. Once all bugs are eliminated, this procedure is complete.

We have thus far shown some procedures for finding the basic causes of bugs. However, we can check them using intermediate output values instead of the final results of a total system. This is regarded as a method of subdividing a total system into subsystems.

4.7. MT and MTS Methods

MT (Mahalanobis–Taguchi) Method

We cannot describe human health with a single measurement. The MT method represents human health by integrating measuring characteristics. We can apply this method to a medical examination or patient monitoring. The concept of distance expressing interrelationships among multidimensional information was introduced by Mahalanobis. However, since his method calculates a distance for each population (group) of samples, it is regarded as an application of deviation value in a one-dimensional space to a multidimensional space. Taguchi proposes introducing distance in a multidimensional space by prescribing the Mahalanobis distance only in the unit cluster of a population and defining variable distance in a population in consideration of variable interrelationships among items.

The MTS method also defines a group of items close to their average as a unit cluster in such a way that we can use it to diagnose or monitor a corporation. In fact, both the MT and MTS methods have started to be applied to various fields because they are outstanding methods of pattern recognition. What is most important in utilizing multidimensional information is to establish a fundamental database. That is, we should consider what types of items to select or which groups to collect to form the database. These issues should be determined by persons expert in a specialized field.

In this section we detail a procedure for streamlining a medical checkup or clinical examination using a database for a group of healthy people (referred to subsequently as "normal" people). Suppose that the total number of items used in the database is k. Using data for a group of normal people—for example, the data of people who are examined and found to be in good health in any year after annual medical checkups for three years in a row (if possible, the data of hundreds of people is preferable)—we create a scale to characterize the group.

As a scale we use the distance measure of P. C. Mahalanobis, an Indian statistician, introduced in his thesis in 1934. In calculating the Mahalanobis distance in a certain group, the group needs to have homogeneous members. In other words, a group consisting of abnormal people should not be considered. If the group contains people with both low and high blood pressure, we should not regard the data as a single distribution.

Indeed, we may consider a group of people suffering only from hepatitis type A as being somewhat homogeneous; however, the group still exhibits a wide deviation. Now, let's look at a group of normal people without hepatitis. For gender and age we include both male and female and all adult age brackets. Male and female are denoted by 0 and 1, respectively. Any item is dealt with as a quantitative measurement in calculation. After regarding 0 and 1 data for male and female as continuous variables, we calculate an average m and standard deviation σ. For all items for normal people, we compute an average value m and standard deviation σ and convert them below. This is called *normalization.*

$$Y = \frac{y - m}{\sigma} \tag{4.153}$$

Suppose that we have n normal people. When we convert $y_1, y_2, \ldots, y_n$ into $Y_1, Y_2, \ldots, Y_n$ using equation (4.153), $Y_1, Y_2, \ldots, Y_n$ has an average of 0 and a standard deviation of 1, leading to easier understanding. Selecting two from k items arbitrarily, and dividing the sum of normalized products by n or calculating covariance, we obtain a correlation coefficient.

After forming a matrix of correlation coefficients between two of k items by calculating its inverse matrix, we compute the following square of D representing the Mahalanobis distance:

$$D^2 = \frac{1}{k}\left(\sum_{ij} a_{ij}Y_iY_j\right) \tag{4.154}$$

Now a_{ij} stands for the (i, j)th element of the inverse matrix. $Y_1, Y_2, \ldots, Y_n$ are converted from $y_1, y_2, \ldots, y_n$ based on the following equations:

$$Y_1 = \frac{y_1 - m_1}{\sigma_1}$$

$$Y_2 = \frac{y_2 - m_2}{\sigma_2}$$

$$\vdots$$

$$Y_k = \frac{y_k - m_k}{\sigma_k} \tag{4.155}$$

In these equations, for k items regarding a group of normal people, set the average of each item to $m_1, m_2, \ldots, m_k$ and the standard deviation to $\sigma_1, \sigma_2, \ldots, \sigma_n$. The data of k items from a person, $y_1, y_2, \ldots, y_k$, are normalized to obtain $Y_1, Y_2, \ldots, Y_k$. What is important here is that "a person" whose identity is unknown in terms of normal or abnormal is an arbitrary person. If the person is normal, D^2 has a value of approximately 1; if not, it is much larger than 1. That is, the Mahalanobis distance D^2 indicates how far the person is from normal people.

For practical purposes, we substitute the following y (representing not the SN ratio but the magnitude of N):

$$y = 10 \log D^2 \tag{4.156}$$

Therefore, if a certain person belongs to a group of normal people, the average of y is 0 dB, and if the person stays far from normal people, y increases because the magnitude of abnormality is enlarged. For example, if y is 20 dB, in terms of D^2 the person is 100 times as far from the normal group as normal people are. In most cases, normal people stay within the range of 0 to 2 dB.

Application of the MT Method to a Medical Diagnosis

Although, as discussed in the preceding section, we enumerate all necessary items in the medical checkup case, we need to beware of selecting an item that is derived from two other items. For example, in considering height, weight, and obesity, we need to narrow the items down to two because we cannot compute an inverse of a matrix consisting of correlation coefficients.

When we calculate the Mahalanobis distance using a database for normal people, we determine the threshold for judging normality by taking into account the following two types of error loss: the loss caused by misjudging a normal person as abnormal and spending time and money to do precise tests; and the loss caused by misjudging an abnormal person as normal and losing the chance of early treatment.

❒ Example

The example shown in Table 4.24 is not a common medical examination but a special medical checkup to find patients with liver dysfunction, studied by Tatsuji

Table 4.24
Physiological examination items

Examination Item	Acronym	Normal Value
Total protein	TP	6.5–7.5 g/dL
Albumin	Alb	3.5–4.5 g./dL
Cholinesterase	ChE	0.60–1.00 ΔpH
Glutamate oxaloacetate transaminase	GOT	2–25 units
Glutamate pyruvate transaminase	GPT	0–22 units
Lactatdehydrogenase	LDH	130–250 units
Alkaline phosphatase	ALP	2.0–10.0 units
γ-Glutamyltranspeptidase	γ-GTP	0–68 units
Lactic dehydrogenase	LAP	120–450 units
Total cholesterol	TCh	140–240 mg/dL
Triglyceride	TG	70–120 g/dL
Phospholipases	PL	150–250 mg/dL
Creatinine	Cr	0.5–1.1 mg/dL
Blood urea nitrogen	BUN	5–23 mg/dL
Uric Acid	UA	2.5–8.0 mg/dL

Kanetaka at Tokyo Teishin Hospital [6]. In addition to the 15 items shown in Table 4.24, age and gender are included.. The total number of items is 17.

By selecting data from 200 people (it is desirable that several hundred people be selected, but only 200 were chosen because of the capacity of a personal computer) diagnosed as being in good health for three years at the annual medical checkup given by Tokyo's Teishin Hospital, the researchers established a database of normal people. The data of normal people who were healthy for two consecutive years may be used. Thus, we can consider the Mahalanobis distance based on the database of 200 people. Some people think that it follows an *F*-distribution with an average of 1 approximate degree of freedom of 17 for the numerator and infinite degrees of freedom for its denominator on the basis of raw data. However, since its distribution type is not important, it is wrong. We compare the Mahalanobis distance with the degree of each patient's dysfunction.

If we use decibel values in place of raw data, the data for a group of normal people are supposed to cluster around 0 dB with a range of a few decibels. To minimize loss by diagnosis error, we should determine a threshold. We show a simple method to detect judgment error next.

Table 4.25 demonstrates a case of misdiagnosis where healthy people with data more than 6 dB away from those of the normal people are judged not normal. For 95 new persons coming to a medical checkup, Kanetaka analyzed actual diagnostic error accurately by using a current diagnosis, a diagnosis using the Mahalanobis distance, and close (precise) examination.

Category 1 in Table 4.25 is considered normal. Category 2 comprises a group of people who have no liver dysfunction but had ingested food or alcohol despite being prohibited from doing so before a medical checkup. Therefore, category 2 should be judged normal, but the current diagnosis inferred that 12 of 13 normal people were abnormal. Indeed, the Mahalanobis method misjudged 9 normal people as being abnormal, but this number is three less than that obtained using the current method.

Category 3 consists of a group of people suffering from slight dysfunctions. Both the current and Mahalanobis methods overlooked one abnormal person. Yet both of them detected 10 of 11 abnormal people correctly.

For category 4, a cluster of 5 people suffering from moderate dysfunctions, both methods found all persons. Since category 2 is a group of normal persons, combining categories 1 and 2 as a group of liver dysfunction $-$, and categories 3 and 4 as a group of liver dysfunction $+$, we summarize diagnostic errors for each method in Table 4.26. For these contingency tables, each discriminability (0: no discriminability, 1: 100% discriminability) is calculated by the following equations (see "Separation System" in Section 4.4 for the theoretical background). A_1's discriminability:

$$\rho_1 = \frac{[(28)(15) - (51)(1)]^2}{(79)(16)(29)(66)} = 0.0563 \tag{4.157}$$

A_2's discriminability:

$$\rho_2 = \frac{(63)(15) - (16)(1)]^2}{(76)(16)(64)(31)} = 0.344 \tag{4.158}$$

Table 4.25
Medical checkup and discriminability

	A_1: Current Method		A_2: Mahalanobis Method		
Category[a]	**Normal**	**Abnormal**	**Normal**	**Abnormal**	**Total**
1	27	39	59	7	66
2	1	12	4	9	13
3	1	10	1	10	11
4	0	5	0	5	5
Total	29	66	64	31	95

[a] 1, Normal; 2, normal but temporarily abnormal due to food and alcohol; 3, slightly abnormal; 4, moderately abnormal.

Table 4.26
2 × 2 Contingency table

A_1: Current method **Diagnosis:** **Liver dysfunction**	**Normal**	**Abnormal**	**Total**
− (Normal)	28	51	79
+ (Abnormal)	1	15	16
Total	29	66	95
A_2: Mahalanobis method **Diagnosis:** **Liver dysfunction**	**Normal**	**Abnormal**	**Total**
− (Normal)	63	16	79
+ (Abnormal)	1	15	16
Total	64	31	95

Each of the equations above consists of a product of peripheral frequencies for the denominator and a squared difference of diagonal products for the numerator. Whereas the discriminability of the current method is 5.63%, that of the Mahalanobis distance is 34.4%, approximately six times as large as that of the current method. This index is the concept of squared correlation coefficient and equivalent to what is called *contribution*.

On the other hand, in expressing discriminability, the following SN ratios are regarded as more rational. A_1's SN ratio:

$$\eta_1 = 10 \log \frac{\rho_1}{1 - \rho_1} = 10 \log \frac{0.0563}{0.9437} = -12.2 \text{ dB} \tag{4.159}$$

A_2's SN ratio:

$$\eta_2 = 10 \log \frac{0.344}{0.656} = -2.8 \text{ dB} \tag{4.160}$$

Therefore, a medical examination using the Mahalanobis distance is better by 9.7 dB or 8.7 times than the current item-by-item examination. According to Table 4.25, both methods have identical discriminability of abnormal people. However, the current method diagnosed 51 of 79 normal people to be or possibly be abnormal, thereby causing a futile close examination. whereas the Mahalanobis method judges only 16 of 79 to be abnormal. That is, the latter enables us to eliminate such a wasteful checkup for 35 normal people.

In calculating the SN ratio, we should use the dynamic SN ratio, as explained below.

Design of General Pattern Recognition and Evaluation Procedure

MT METHOD

There is a large gap between science and technology. The former is a quest for truth; the latter does not seek truth but invents a means of attaining an objective function. In mechanical, electronic, and chemical engineering, technical means are described in a systematic manner for each objective function. For the same objective function, there are various types of means, which do not exist in nature.

A product planning department considers types of functions. The term *product planning*, including hardware and software, means to plan products to be released in the marketplace, most of which are related to active functions. However, there are quite a few passive functions, such as inspection, diagnosis, and prediction. In evaluating an active function, we focus on the magnitude of deviation from its ideal function. Using the SN ratio, we do a functional evaluation that indicates the magnitude of error.

In contrast, a passive function generally lies in the multidimensional world, including time variables. Although our objective is obviously pattern recognition in the multidimensional world, we need a brand-new method of summarizing multidimensional data. To summarize multidimensional data, quality engineering (more properly, the Taguchi method) gives the following paradigms:

- *Procedure 1:* Select items, including time-series items. Up to several thousand items can be analyzed by a computer.
- *Procedure 2:* Select zero-point and unit values and base space to determine a pattern.
- *Procedure 3:* After summarizing measurement items only from the base space, we formulate the equation to calculate the Mahalanobis distance. The Mahalanobis space determines the origin of multidimensional measurement and unit value.
- *Procedure 4:* For an object that does not belong to the base space, we compute Mahalanobis distances and assess their accuracy using an SN ratio.
- *Procedure 5:* Considering cost and the SN ratio, we sort out necessary items for optimal diagnosis and proper prediction method.

Among these specialists are responsible for procedures 1 and 2 and quality engineering takes care of procedures 3, 4, and 5. In quality engineering, any quality problem is attributed to functional variability. That is, functional improvement will lead to solving any type of quality problem. Therefore, we determine that we should improve functionality by changing various types of design conditions (when dealing with multiple variables, and selecting items and the base space).

Since the Mahalanobis distance (variance) rests on the calculation of all items selected, we determine the following control factors for each item or for each group of items and, in most cases, allocate them to a two-level orthogonal array when we study how to sort items:

- *First factor level.* Items selected (or groups of items selected) are used.
- *Second factor level.* Items selected (or groups of items selected) are not used.

We do not need to sort out items regarded as essential. Now suppose that the number of items (groups of items) to be studied is l and an appropriate two-level orthogonal array is L_N. When $l = 30$, L_{32} is used, and when $l = 100$, L_{108} or L_{124} is selected.

Once control factors are assigned to L_N, we formulate an equation to calculate the Mahalanobis distance by using selected items because each experimental condition in the orthogonal array shows the assignment of items. Following procedure 4, we compute the SN ratio, and for each SN ratio for each experiment in orthogonal array L_N, we calculate the control factor effect. If certain items chosen contribute negatively or little to improving the SN ratio, we should select an optimal condition by excluding the items. This is a procedure of sorting out items used for the Mahalanobis distance.

MTS METHOD

Although an orthogonalizing technique exists that uses principal components when we normalize data in a multidimensional space, it is often unrelated to economy and useless because it is based too heavily on mathematical background. Now we introduce a new procedure for orthogonalizing data in a multidimensional space, which at the same time reflects on the researchers' objective.

We select $X_1, X_2, \ldots, X_k$ as k-dimensional variables and define the following as n groups of data in the Mahalanobis space:

$$\begin{array}{l} X_{11}, X_{12}, \ldots, X_{1n} \\ X_{21}, X_{22}, \ldots, X_{2n} \\ \vdots \\ X_{k1}, X_{k2}, \ldots, X_{kn} \end{array}$$

All of the data above are normalized. That is, n data, $X_1, X_2, \ldots, X_n$ have a mean of zero and a variance of 1. $X_1, X_2, \ldots, X_k$ represent the order of cost or priority, which is an important step and should be determined by engineers. In lieu of X_1, $X_2, \ldots, X_k$, we introduce new variables, $x_1, x_2, \ldots, x_k$ that are mutually orthogonal:

$$x_1 = X_1 \tag{4.161}$$

x_1 is equal to X_1 and has a mean of zero and a variance of 1. Variable x_2 is the part of variable X_2, that remains after removing the part related to, except for the part involving variable x_1 (or X_1). We consider the following equation, where X_2 is expressed by x_1:

$$X_2 = b_{21}x_1 \tag{4.162}$$

In this case b_{12} is not only the regression coefficient but also the correlation coefficient. Then the remaining part of X_2, excluding the part related to x_1 (regression part) or the part independent of x_1, is

$$x_2 = X_2 - b_{21}x_1 \tag{4.163}$$

Orthogonality of x_1 and x_2 indicates that a sum of products of x_{1j} and x_{2j} amounts to zero in the base space:

$$\sum_{j=1}^{n} x_{1j}x_{2j} = \sum_{j=1}^{n} X_{1j}(X_{2j} - b_{21}x_{1j}) = 0 \tag{4.164}$$

Thus, x_1 and x_2 become orthogonal. On the other hand, whereas x_1's variance is 1, x_2's variance σ_2^2, called *residual contribution*, is calculated by the equation

$$\sigma_2^2 = 1 - b_{21}^2 \tag{4.165}$$

The reason is that if we compute the mean square of the residuals, we obtain the following result:

$$\begin{aligned}\frac{1}{n}\sum(X_{2j} - b_{21}x_{1j})^2 &= \frac{1}{n}\sum X_{2j}^2 - \frac{2}{n}b_{21}\sum X_{2j}x_{ij} + \frac{1}{n}(-b_{21})^2\sum x_{1j}^2 \\ &= 1 - \frac{2n}{n}b_{2l}^2 + \frac{n}{n}b_{2l}^2 \\ &= 1 - b_{2l}^2\end{aligned} \tag{4.166}$$

We express the third variable, X_3, with x_1 and x_2:

$$X_3 = b_{31}x_1 + b_{32}x_2 \tag{4.167}$$

By multiplying both sides of equation (4.167) by x_1, calculating a sum of elements in the base space, and dividing by n, we obtain

$$b_{31} = \frac{1}{nV_1}\sum X_{3j}x_{1j} \tag{4.168}$$

$$V_1 = \frac{1}{n}\sum x_{1j}^2 \tag{4.169}$$

Similarly, by multiplying both sides of equation (4.167) by x_2, we have b_{32}

$$b_{32} = \frac{1}{nV_2}\sum X_{3j}x_{2j} \tag{4.170}$$

$$V_2 = \frac{1}{n}\sum x_{2j}^2 \tag{4.171}$$

The orthogonal part of x_3 is thus a residual part, so that X_3 cannot be expressed by both X_1 and X_2 or both x_1 and x_2. In sum,

$$x_3 = X_3 - b_{31}x_1 - b_{32}x_2 \tag{4.172}$$

In the Mahalanobis space, x_1, x_2, and x_3 are orthogonal. Since we have already proved the orthogonality of x_1 and x_2, we prove here that of x_1 and x_3 and that of x_2 and x_3. First, considering the orthogonality of x_1 and x_3, we have

$$\begin{aligned}\sum_j x_{1j}(X_{3j} - b_{31}x_{1j} - b_{32}x_{2j}) &= \sum_j x_{1j}X_{3j} - b_{31}\sum x_{1j}^2 \\ &= 0\end{aligned} \tag{4.173}$$

This can be derived from equation (4.167) defining b_{31}. Similarly, the orthogonality of x_2 and x_3 is proved according to equation (4.170), defining b_{32} as

$$\begin{aligned}\sum x_{2j}\,(X_{3j} - b_{31}x_{1j} - b_{32}x_{2j}) &= \sum x_{2j}X_{3j} - b_{32}\sum x_{2j}^2 \\ &= 0 \end{aligned} \tag{4.174}$$

For the remaining variables, we can proceed with similar calculations. The variables normalized in the preceding section can be rewritten as follows:

$$\begin{aligned} x_1 &= X_1 \\ x_2 &= X_2 - b_{21}x_1 \\ x_3 &= X_3 - b_{31}x_1 - b_{32}x_2 \\ &\vdots \\ x_k &= X_k - b_{k1}x_1 - b_{2k}x_2 - \cdots - b_{k(k-1)}x_{k-1} \end{aligned} \tag{4.175}$$

Each of the orthogonalized variables $x_2, \ldots, x_k$ does not have a variance of 1. In an actual case we should calculate a variance right after computing n groups of variables, $x_1, x_2, \ldots, x_n$ in the Mahalanobis space. As for degrees of freedom in calculating a variance, we can regard n for x_1, $n - 1$ for x_2, $n - 2$ for $x_3, \ldots, n - k + 1$ for x_k. Instead, we can select n degrees of freedom for all because $n \gg k$ quite often.

$$\begin{aligned} V_1 &= \frac{1}{n}\,(x_{11}^2 + x_{12}^2 + \cdots + x_{1n}^2) \\ V_2 &= \frac{1}{n-1}\,(x_{21}^2 + x_{22}^2 + \cdots + x_{2n}^2) \\ &\vdots \\ V_k &= \frac{1}{n-k+1}\,(x_{k1}^2 + x_{k2}^2 + \cdots + x_{kn}^2) \end{aligned} \tag{4.176}$$

Now setting normalized, orthogonal variables to $y_1, y_2, \ldots, y_k$, we obtain

$$\begin{aligned} y_1 &= \frac{x_1}{\sqrt{V_1}} \\ y_2 &= \frac{x_2}{\sqrt{V_2}} \\ &\vdots \\ y_k &= \frac{x_k}{\sqrt{V_k}} \end{aligned} \tag{4.177}$$

All of the normalized $y_1, y_2, \ldots, y_k$ are orthogonal and have a variance of 1. The database completed in the end contains m and σ as the mean and standard deviation of initial observations, b_{21}, V_2, b_{31}, b_{32}, $V_2, \ldots, b_{k1}, b_{k2}, \ldots, b_{k(k-1)}$, V_k as the coefficients and variances for normalization. Since we select n groups of data, there are k means of $m_1, m_2, \ldots, m_k$, k standard deviations of $\sigma_1, \sigma_2, \ldots, \sigma_k$, $(k-1)k/2$ coefficients, and k variances in this case. Thus, the number of necessary memory items is as follows:

$$k + k + \frac{(k-1)k}{2} + k = \frac{k\,(k+5)}{2} \qquad (4.178)$$

The correlation matrix in the normalized Mahalanobis space turns out to be an identity matrix. Then the correlation matrix R is expressed as

$$R = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & 1 \end{pmatrix} \qquad (4.179)$$

Therefore, the inverse matrix of R, A is also an identity matrix:

$$A = R^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & 1 \end{pmatrix} \qquad (4.180)$$

Using these results, we can calculate the Mahalanobis distance D^2 as

$$D^2 = \frac{1}{k}\,(y_1^2 + y_2^2 + \cdots + y_k^2) \qquad (4.181)$$

Now, setting measured data already subtracted by m and divided by σ to $X_1, X_2, \ldots, X_k$, we compute primarily the following $x_1, x_2, \ldots, x_k$:

$$\begin{aligned} x_1 &= X_1 \\ x_2 &= X_2 - b_{21}x_1 \\ x_3 &= X_3 - b_{31}x_1 - b_{32}x_2 \\ &\vdots \\ x_k &= X_k - b_{k1}x_1 - \cdots - b_{k(k-1)} \end{aligned} \qquad (4.182)$$

Thus, $y_1, y_2, \ldots, y_k$ are calculated as

$$\begin{aligned} y_1 &= x_1 \\ y_2 &= \frac{x_2}{\sqrt{V_2}} \\ y_3 &= \frac{x_3}{\sqrt{V_3}} \\ &\vdots \\ y_k &= \frac{x_k}{\sqrt{V_k}} \end{aligned} \qquad (4.183)$$

When for an arbitrary object we calculate $y_1, y_2, \ldots, y_k$ and D^2 in equation (4.181), if a certain variable belongs to the Mahalanobis space, D^2 is supposed to take a value of 1, as discussed earlier. Otherwise, D^2 becomes much larger than 1 in most cases.

Once the normalized orthogonal variables $y_1, y_2, \ldots, y_k$ are computed, the next step is the selection of items.

A_1: Only y_1 is used

A_2: y_1 and y_2 are used

⋮

A_k: $y_1, y_2, \ldots, y_k$ are used

Now suppose that values of signal factor levels are known. Here we do not explain a case of handling unknown values. Although some signals belong to the base space, l levels of a signal that is not included in the base space are normally used. We set the levels to $M_1, M_2, \ldots, M_l$. Although l can only be 3, we should choose as large an l value as possible to calculate errors.

In the case of A_k, or the case where all items are used, after calculating the Mahalanobis distances for $M_1, M_2, \ldots, M_l$ by taking the square root of each, we create Table 4.27. What is important here is not D^2 but D per se. As a next step, we compute dynamic SN ratios. Although we show the case where all items are used, we can create a table similar to Table 4.27 and calculate the SN ratio for other cases, such as the case when partial items are assigned to an orthogonal array.

Based on Table 4.27, we can compute the SN ratio η as follows:

Total variation:

$$S_T = D_1^2 + D_2^2 + \cdots + D_1^2 \qquad (f = l) \tag{4.184}$$

Variation of proportional terms:

$$S_\beta = \frac{(M_1D_1 + M_2D_2 + \cdots + M_lD_l)^2}{r} \qquad (f = 1) \tag{4.185}$$

Effective divider:

$$r = M_1^2 + M_2^2 + \cdots + M_l^2 \tag{4.186}$$

Error variation:

$$S_e = S_T - S_\beta \qquad (f = l - 1) \tag{4.187}$$

Table 4.27
Signal values and Mahalanobis distance

Signal-level value	M_1	M_2	⋯	M_k
Mahalanobis distance	D_1	D_2	⋯	D_l

SN ratio:

$$\eta = 10 \log \frac{(1/r)(S_\beta - V_e)}{V_e} \tag{4.188}$$

On the other hand, for the calibration equation, we calculate

$$\beta = \frac{M_1D_1 + M_2D_2 + \cdots + M_lD_l}{r} \tag{4.189}$$

and we estimate

$$M = \frac{D}{\beta} \tag{4.190}$$

In addition, for $A_1, A_2, \ldots, A_{k-1}$, we need to compute SN ratios, $\eta_1, \eta_2, \ldots, \eta_{k-1}$. Table 4.28 summarizes the result. According to the table, we determine the number of items by balancing SN ratio and cost. The cost is not the calculation cost but the measurement cost for items. We do not explain here the use of loss functions to select items.

Summary of Partial MD Groups: Countermeasure for Collinearity

To summarize the distances calculated from each subset of distances to solve collinearity problems, this approach can be widely applied. The reason is that we can select types of scale to use, as what is important is to select a scale that expresses patients' conditions accurately no matter what correlation we have in the Mahalanobis space. The point is to be consistent with patients' conditions diagnosed by doctors.

Now let's go through a new construction method. For example, suppose that we have data for 0 and 1 in a 64 × 64 grid for computer recognition of handwriting. If we use 0 and 1 directly, 4096 elements exist. Thus, the Mahalanobis space formed by a unit set (suppose that we are dealing with data for n persons in terms of whether a computer can recognize a character "A" as "A": for example, data from 200 sets in total if 50 people write a character of four items) has a 4096 × 4096 matrix. This takes too long to handle using current computer capability.

How to leave out character information is an issue of information system design. In fact, a technique for substituting only 128 data items consisting of 64 column sums and 64 row sums for all data in a 64 × 64 grid has already been proposed. Since two sums of 64 column sums and 64 row sums are identical, they have collinearity. Therefore, we cannot create a 128 × 128 unit space by using the relationship 64 + 64 = 128.

In another method, we first create a unit space using 64 row sums and introduce the Mahalanobis distance, then create a unit space using 64 column sums and

Table 4.28
Signal values and Mahalanobis distances

Number of items	1	2	3	⋯	k
SN ratio	η_1	η_2	η_3	⋯	η_k
Cost	C_1	C_2	C_3	⋯	C_k

calculate the Mahalanobis distance. For a few alphabets similar to "B," we calculate Mahalanobis distances. If 10 persons, N_1. N_2, ... , N_{10} write three letters similar to "B," we can obtain 30 signals. After each of the 10 persons writes the three letters "D," "E," and "R," we compute the Mahalanobis distances for all signals from a unit space of "B." This is shown in Table 4.29.

We calculate a discriminability SN ratio according to this table. Because we do not know the true difference among M's, we compute SN ratios for unknown true values as follows:

Total variation:

$$S_T = D_{11}^2 + D_{12}^2 + \cdots + D_{3,10}^2 \qquad (f = 30) \tag{4.191}$$

Signal effect:

$$S_M = \frac{D_1^2 + D_2^2 + D_3^2}{10} \qquad (f = 3) \tag{4.192}$$

$$V_M = \frac{S_M}{3} \tag{4.193}$$

$$S_e = S_T - S_M \qquad (f = 27) \tag{4.194}$$

$$V_e = \frac{S_e}{27} \tag{4.195}$$

By calculating row sums and column sums separately, we compute the following SN ratio η using averages of V_M and V_e:

$$\eta = 10 \log \frac{\frac{1}{10}(V_M - V_e)}{V_e} \tag{4.196}$$

The error variance, which is a reciprocal of the antilog value, can be computed as

$$\sigma^2 = 10^{-1.1\eta} \tag{4.197}$$

The SN ratio for selection of items is calculated by setting the following two levels:

- ❑ *Level 1:* item used
- ❑ *Level 2:* item not used

We allocate them to an orthogonal array. By calculating the SN ratio using equation (4.196), we choose optimal items.

Table 4.29
Mahalanobis distances for signals

Signal	Noise				Total
	N_1	N_2	...	N_{10}	
M_1 (D)	D_{11}	D_{12}	...	$D_{1,10}$	D_1
M_2 (E)	D_{21}	D_{22}	...	$D_{2,10}$	D_2
M_3 (R)	D_{31}	D_{32}	...	$D_{3,10}$	D_3

Although there can be some common items among groups of items, we should be careful not to have collinearity in one group. Indeed, it seems somewhat unusual that common items are included in several different groups, even though each common item should be emphasized; however, this situation is considered not so unreasonable because the conventional single correlation has neglected other, less related items completely.

A key point here is that we judge whether the summarized Mahalanobis distance corresponds to the case of using the SN ratio calculated by Mahalanobis distances for items not included in the unit space. Although we can create Mahalanobis spaces by any procedure, we can judge them only by using SN ratios.

4.8. On-line Quality Engineering

None of us cares about causes of errors when we set a clock. We sometimes calibrate the time by comparing it with a time signal from radio or TV. This is the case with almost all characteristics in the production process. Let's say that according to company S, the cost of comparing the time with the time signal is approximately $1.50. In addition, assuming that calibration takes 5 seconds, it will cost around 10 cents. A quality control system in the production process is responsible for handling eventual clock errors, given $1.50 for comparison of the time with the time signal and 10 cents for calibration. If we predict such errors, we can determine optimal checkup intervals or calibration limits using loss functions.

A technical procedure for designing optimal checkup intervals and adjustment limits is called *on-line quality engineering* or *quality control in process*. In on-line quality engineering, we should clarify three system elements used to design a system: process, checkup procedure (comparison with a standard such as a time signal and generally called *measurement*), and correction method (in the case of calibration of a measuring machine called *calibration*, and in the case of process called *correction* or *adjustment*). In this section we discuss a quality control system using a simple example.

❒ Example

Although automobile keys are generally produced as a set of four identical keys, this production system is regarded as dynamic because each set has a different dimension and shape. Each key has approximately 9 to 11 cuts, and each cut has several different dimensions. If there are four different dimensions with a 0.5-mm step, a key with 10 cuts has $4^{10} = 1{,}048{,}576$ variations.

In actuality, each key is produced such that it has a few different-dimension cuts. Then the number of key types in the market will be approximately 10,000. Each key set is produced based on the dimensions indicated by a computer. By using a master key we can check whether a certain key is produced, as indicated by the computer. Additionally, to manage a machine, we can examine particularshaped keys produced by such machines at certain intervals. Now we consider a case where we conduct process quality control by measuring differences between dimensions of cuts of the last key in one set and those shown by a computer.

For the sake of convenience, the standard of key dimensions is ± 30 μm for each target value. Whether or not this ± 30 μm tolerance is rational depends on whether the safety factor of 8 is appropriate for the function limit of ± 250 μm. (In this case, since the height of a cut is designed using a 500-μm step, if a certain dimension deviates by 250 μm more or less than its design dimension, it can be taken for other designs. Thus, the function limit is 250 μm. In this case, is a safety factor of 8 appropriate? If we suppose the price of a key set shipped from a plant to be \$1.90, the eventual loss in the market when a key does not work becomes 64 times as great as the price, or \$1.20. We consider the tolerance appropriate; that is, when the malfunction of the key costs \$120, ± 30 μm makes sense.

A manufacturing department's task is not to determine tolerances for a product but to produce a product with dimensions that lie within a range of tolerances. To achieve this, how should the manufacturing department design a quality control system? For example, in an inspection room, once an hour we check the dimensions of a key shortly after it has been machined. When we inspect every n products produced, we call the number n the *inspection interval* (also called *diagnosis, checkup,* or *measurement interval*). If we produce 300 sets in an hour, the current inspection interval N_0 is 300 sets. Now the inspection cost B when we pick up a product in the production process and check it in the inspection room (including labor cost) amounts to \$12. In addition, the loss A for one defective set is already assumed to be \$1.90. On the other hand, the number of key sets produced while the key selected is inspected for 10 minutes is called the *time lag of inspection*. In this case we regard this as 50 sets. At present, we are controlling the production process within a range of ± 20 μm, which is two-thirds of the current tolerance, $\Delta = 30$ μm. Then the current adjustment limit D_0 is 20 μm. Currently, one of 19,560 products goes beyond the adjustment limit of ± 20 (μm). Since we stop the production line and take countermeasures such as an adjustment or tool change when a product exceeds the limit, we need an adjustment cost C of \$58. In sum, we enumerate all preconditions for designing a process control system:

- ❑ *Dimensional tolerance:* $\Delta = 30$ μm
- ❑ *Loss of defective product:* $A =$ \$1.90 per set
- ❑ *Inspection cost:* $B =$ \$12
- ❑ *Time lag between inspection and judgment:* $l = 50$ sets
- ❑ *Adjustment cost:* $C =$ \$58
- ❑ *Current inspection interval:* $n_0 = 300$ sets
- ❑ *Current adjustment limit:* $D_0 = 20$ μm
- ❑ *Observed mean adjustment interval:* $u_0 = 19{,}560$ sets

When we inspect a set in an hour and control the production process with the adjustment limit $D_0 = 20$ μm, the only observation is how many inspections happen from one to another adjustment. In this case, this value, u_0, indicating process stability, is 19,560.

Design of Feedback Control System

To design all process control systems in an economical manner is the most crucial task for managers. The cost of the current control system for one set is

$$\text{control cost} = \frac{B}{n_0} + \frac{C}{u_0} \tag{4.198}$$

In addition, we need a monetary evaluation of the quality level according to dimensional variability. When the adjustment limit is $D_0 = \pm 20$ μm, the dimensions are regarded as being distributed uniformly within the limit because they are not controlled in the range. The variance is as follows:

$$\frac{D_0^2}{3} \tag{4.199}$$

On the other hand, the magnitude of dispersion for an inspection interval n_0 and an inspection time lag of l is

$$\left(\frac{n_0 + 1}{2} + l\right)\frac{D_0^2}{u_0} \tag{4.200}$$

Consequently, we obtain the following loss function:

$$\frac{A}{\Delta^2}\left[\frac{D_0^2}{3} + \left(\frac{n_0 + 1}{2} + l\right)\frac{D_0^2}{u_0}\right] \tag{4.201}$$

Adding equations (4.198) and (4.201) to calculate the following economic loss L_0 for the current control system, we have 33.32 cents per product:

$$\begin{aligned} L_0 &= \frac{B_0}{n_0} + \frac{C}{u_0} + \frac{A}{\Delta^2}\left[\frac{D_0^2}{3}\left(\frac{n_0 + 1}{2} + l\right)\frac{D_0^2}{u_0}\right] \\ &= \frac{12}{300} + \frac{58}{19{,}560} + \frac{1.90}{30^2}\left[\frac{20^2}{3} + \left(\frac{301}{2} + 50\right)\frac{20^2}{19{,}560}\right] \\ &= 0.04 + 000.30 + 0.2815 + 0.87 \\ &= 33.32 \text{ cents} \end{aligned} \tag{4.202}$$

Assuming that the annual operation time is 1600 hours, we can see that in the current system the following amount of money is spent annually to control quality:

$$(33.32)(300)(1600) \approx \$160{,}000 \tag{4.203}$$

If an error σ_m accompanies each measurement, another loss of measurement is added to the loss above:

$$\frac{A}{\Delta^2}\,\sigma_m^2 \tag{4.204}$$

It is an optimal control system that improves the loss in equation (4.203). In short, it is equivalent to determining an optimal inspection interval n and adjustment limit D, both of which are calculated by the following formulas:

$$n = \sqrt{\frac{2u_0B}{A}\frac{\Delta}{D_0}} \tag{4.205}$$

$$= \sqrt{\frac{(2)(19{,}560)(12)}{1.90}}\left(\frac{30}{20}\right)$$

$$= 745$$

$$\approx 600 \text{ (twice an hour)} \tag{4.206}$$

$$D = \left(\frac{3C}{A} \times \frac{D_0^2}{u_0}\Delta^2\right)^{1/2} \tag{4.207}$$

$$= \left[\frac{(3)(58)}{1.90}\left(\frac{20^2}{19{,}560}\right)\Delta^2\right]^{1/4}$$

$$= 6.4 \tag{4.208}$$

$$\approx 7.0\ \mu\text{m} \tag{4.209}$$

Then, by setting the optimal measurement interval to 600 sets and adjustment limit to ± 7.0 μm, we can reduce the loss L:

$$L = \frac{B}{n} + \frac{C}{u} + \frac{A}{\Delta^2}\left[\frac{D^2}{32} + \left(\frac{n+1}{2} + l\right)\frac{D^2}{u}\right] \tag{4.210}$$

Using this equation, we estimate the mean adjustment interval u:

$$u = u_0\frac{D^2}{D_0^2}$$

$$= (19{,}560)\left(\frac{7^2}{20^2}\right)$$

$$= 2396 \tag{4.211}$$

Therefore, we need to change the adjustment interval from the current level of 65 hours to 8 hours. However, the total loss L decreases as follows:

$$L = \frac{12}{600} + \frac{58}{2396} + \frac{1.80}{30^2}\left[\frac{7^2}{3} + \left(\frac{601}{2} + 50\right)\frac{7^2}{2396}\right]$$

$$= 0.02 + 0.0242 + 3.45 + 1.51$$

$$= 9.38 \text{ cents} \tag{4.212}$$

This is reduced from the current level by

$$33.32 - 9.38 = 23.94 \text{ cents} \tag{4.213}$$

On a yearly basis, we can expect an improvement of

$$(23.94)(300)(1600) = \$114{,}912 \tag{4.214}$$

Management in Manufacturing

The procedure described in the preceding section is a technique of solving a balancing equation of checkup and adjustment costs, and necessary quality level, leading finally to the optimal allocation of operators in a production plant. Now, given that checkup and adjustment take 10 and 30 minutes, respectively, the work hours required in an 8-hour shift at present are:

(10 min) × (daily no. inspections) + (30 min) × (daily no. adjustments)

$$= (10)\,\frac{(8)(300)}{n} + (30)\,\frac{(3)(300)}{u} \tag{4.215}$$

$$= (10)\left(\frac{2400}{600}\right) + (30)\left(\frac{2400}{2396}\right)$$

$$= 40 + 30.1 = 70 \text{ minutes} \tag{4.216}$$

Assuming that one shift has a duration of 8 hours, the following the number of workers are required:

$$\frac{70}{(8)(60)} = 0.146 \text{ worker} \tag{4.217}$$

If there are at most six processes,

$$(0.146)(6) = 0.88 \text{ worker} \tag{4.218}$$

is needed, which implies that one operator is sufficient. As compared to the following workers required currently, we have

$$(10)\,\frac{(8)(300)}{300} + (30)\,\frac{(3)(300)}{19{,}560} = 83.6 \text{ minutes} \tag{4.219}$$

so we can reduce the number of workers by only 0.028 in this process. On the other hand, to compute the process capability index C_p, we estimate the current standard deviation σ_0 using the standard deviation of a measurement error σ_m:

$$\sigma_0 = \sqrt{\frac{D_0^2}{3} + \left(\frac{n_0 + 1}{2} + l\right)\frac{D_0^2}{u_0} + \sigma_m^2} \tag{4.220}$$

Now setting $\sigma_m = 2\ \mu\text{m}$, we obtain σ_0:

$$\sigma_0 = \sqrt{\frac{20^2}{3} + \left(\frac{301}{2} + 50\right)\frac{20^2}{19560} + 2^2}$$

$$= 11.9\ \mu\text{m} \tag{4.221}$$

Thus, the current process capability index C_p is computed as follows:

$$C_p = \frac{2\Delta}{6\sigma_0} = \frac{(2)(30)}{(6)(11.9)} = 0.84 \tag{4.222}$$

The standard deviation in the optimal feedback system is

$$\sigma = \sqrt{\frac{7^2}{3} + \left(\frac{601}{2} + 50\right)\frac{7^2}{2396} + 2^2} = 5.24\ \mu\text{m} \tag{4.223}$$

Then we cannot only reduce the required workforce by 0.028 worker (0.84 times) but can also enhance the process capability index C_p from the current level to

$$C_p = \frac{(30)(2)}{(6)(5.24)} = 1.91 \tag{4.224}$$

Manufacturing Strategy: Balance of Production Measurements and Quality Control System

The most essential strategy planned by manufacturing managers is not to balance quality and cost but to improve productivity. The difference in profitability between corporations is primarily the difference in production speed. Quite a few managers insist that we should continue to produce only a certain volume because even if the production speed is doubled and the production volume is doubled, the increased volume tends to remain unsold in most cases. Nevertheless, we believe that a key point is to develop a technology to double the production speed to prepare for such a demand. As an extreme idea, after doubling the production speed, we can stop after half a day. In this case, an R&D department could offer new jobs to idle workers, or after lowering the retail price by two-thirds, could double sales.

Unless a company sells a product at a retail price several times as high as a plant price (UMC or production cost), it cannot survive. This is because labor and running expenses in sales, administration, and R&D departments account for a few percent of its total cost. In other words, the retail price needs to include the total running cost and profit. When a product is sold at a price three times as high as a plant price, if the production volume is doubled, the plant price, excluding material cost, will be cut in half. Therefore, if the material cost accounts for 30% of the plant price, in cases of double production speed, the retail price can be reduced to an amount 1.65 times as great as the plant cost:

$$[0.3 + \tfrac{1}{2}(0.7)] + \tfrac{1}{2}(2) = 1.65 \text{ times} \tag{4.225}$$

As compared to 3 times, 1.65 times implies that we can lower the retail price 0.55-fold. This estimation is grounded on the assumption that if we sell a product at 45% off the original price, the sales volume doubles. This holds true when we offer new jobs to half the workers, with the sales volume remaining at the same level. Now, provided that the mean adjustment interval decreases by one-fourth due to increased variability after the production speed is raised, how does the cost eventually change? When frequency of machine breakdowns quadruples, u decreases from its current level of 2396 to 599, one-fourth of 2396. Therefore, we alter the current levels of u_0 and D_0 to the values $u_0 = 599$ and $D_0 = 7$ μm. By taking these into account, we consider the loss function. The cost is cut by 0.6-fold and the production volume is doubled. In general, as production conditions change, the optimal inspection interval and adjustment limit also change. Thus, we need to recalculate n and D here. Substituting 0.6A for A in the formula, we obtain

$$
\begin{aligned}
n &= \sqrt{\frac{2u_0B}{0.6A}\,\frac{\Delta}{D_0}} \\
&= \sqrt{\frac{(2)(599)(12)}{0.6 \times 190}\left(\frac{30}{7}\right)} \\
&= 481 \longrightarrow 600 \quad \text{(once in an hour)}
\end{aligned}
\tag{4.226}
$$

$$
\begin{aligned}
D &= \left[\frac{(3)(58)}{(0.6)(1.90)}\left(\frac{7^2}{599}\right)(30^2)\right]^{1/4} \\
&= 10.3 \longrightarrow 10.00
\end{aligned}
\tag{4.227}
$$

On the other hand, the mean adjustment interval is as follows:

$$
u = (599)\left(\frac{10^2}{7^2}\right) = 1222 \tag{4.228}
$$

Hence, the loss function is computed as follows:

$$
\begin{aligned}
L &= \frac{B}{n} + \frac{C}{u} + \frac{0.6A}{30^2}\left[\frac{D^2}{3} + \left(\frac{n+1}{2} + 2l\right)\frac{D^2}{u}\right] \\
&= \frac{12}{600} + \frac{58}{1222} + \frac{(0.6)(1.90)}{30^2}\left[\frac{10^2}{3} + \left(\frac{601}{2} + 100\right)\frac{10^2}{1222}\right] \\
&= 0.02 + 0.0475 + 4.22 + 4.15 \\
&= 15.12 \text{ cents}
\end{aligned}
\tag{4.229}
$$

As a consequence, the quality control cost increases by 5.74 cents (= 15.12 − 9.38). However, the total loss including cost is

$$
\$1.90 + 9.38 \text{ cents} = \$1.99 \tag{4.230}
$$

$$
(0.6)(190 \text{ cents}) + 15.12 \text{ cents} = \$1.29 \tag{4.231}
$$

Suppose that the production volume remains the same, (300)(1600) = 480,000 sets. We can save the following amount of money on an annual basis:

$$(1.99 - 1.29)(48) = \$33{,}600{,}000 \qquad (4.232)$$

References

1. Genichi Taguchi, 1987. *System of Experimental Design*. Dearborn, Michigan: Unipub/American Supplier Institute.
2. Genichi Taguchi, 1984. *Reliability Design Case Studies for New Product Development*. Tokyo: Japanese Standards Association.
3. Genichi Taguchi et al., 1992. *Technology Development for Electronic and Electric Industries*. Quality Engineering Application Series. Tokyo: Japanese Standards Association.
4. Measurement Management Simplification Study Committee, 1984, *Parameter Design for New Product Development*. Japanese Standards Association.
5. Genichi Taguchi et al., 1989. *Quality Engineering Series*, Vol. 2. Tokyo: Japanese Standards Association.
6. Tatsuji Kenetaka, 1987. *An application of Mahalanobis distance. Standardization and Quality Control*, Vol. 40, No. 10.