6.5.1 OVERVIEW OF RAW MATERIAL MIXING PROBLEMS; PURPOSE OF RAW MATERIAL MIXING

Here, we take a look at a case in which T Method-1 was used to deduce the tensile strength of a product from the mixing ratio of raw materials. As shown in Figure 6.25, the product was made from a mixture of five raw materials and two additives in specific mixture ratios, which were blended and submitted to the manufacturing process. The purpose of this procedure was to establish an estimation of the tensile strength of the product (abbreviated as "strength" hereafter) based on the raw material blending ratio.

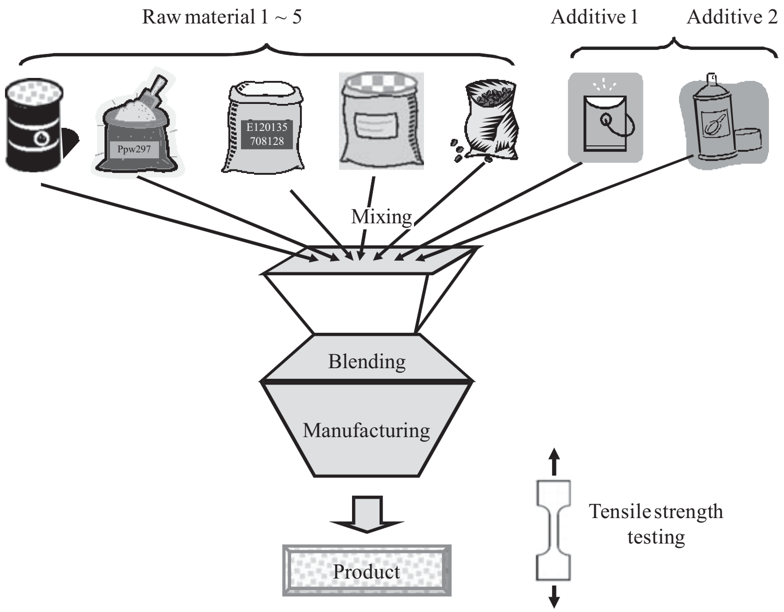


Figure 6.25. Raw material mixing and processing.

Table 6.7 shows the data for mixture ratios of raw materials and strength. The table displays the relative weight ratios (%) of the five raw materials and two additives in relation to the overall weight of 100 (%). In this example, the individual weights of raw materials 1, 2, 4, and 5 and, additives 1 and 2 were independently decided; later, raw material 3 was assigned an appropriate value to make the overall weight 100%. The output value represented the strength in terms of Mpa.

6.5.2 DEFINITION OF THE UNIT SPACE

From among the 10 sets of data in Table 6.7, we designate two, Nos. 5 and 6, the respective strengths of which figure in the vicinity of average values, as Unit Space data, and find the

Table 6.7. Raw material mixing ratio and strength

	Raw materials					Additives		
No.	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**Strength**
1	34.27	7.10	20.08	24.30	9.48	1.17	3.60	49.77
2	26.78	21.71	15.23	23.84	7.00	1.74	3.70	53.73
3	17.01	26.04	19.65	23.16	9.41	1.12	3.60	54.10
4	23.77	22.25	15.40	25.67	7.00	2.21	3.70	54.29
5	22.11	21.71	19.91	23.84	7.00	1.74	3.70	56.27
6	22.14	30.49	11.15	23.88	7.00	1.74	3.60	56.45
7	22.11	21.71	19.91	23.84	7.00	1.74	3.70	59.14
8	20.81	21.05	19.25	26.56	7.00	1.63	3.69	59.89
9	12.18	31.64	19.91	23.84	7.00	1.74	3.70	60.59
10	19.66	23.15	21.35	22.75	7.00	2.37	3.71	61.51

avg=56.574

averages of the two pieces of data with respect to each item and strength. The result of this is displayed in Table 6.8.

Table 6.8. Members of the unit space and averages of items and outputs

	Raw materials					Additives		
No.	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**Strength**
5	22.11	21.71	19.91	23.84	7.00	1.74	3.70	56.27
6	22.14	30.49	11.15	23.88	7.00	1.74	3.60	56.45
Average	22.13	26.1	15.53	23.86	7.00	1.74	3.65	56.36

6.5.3 DEFINITION OF SIGNAL DATA

Next, from all 10 sets of data, we chose as Signal Data the remaining eight pieces of data excluding the two items of data selected for the Unit Space. Table 6.9 reflects how this works out. The Signal Data is used for the computation of proportional coefficient β and SN ratio η, which will be discussed in Subsection 6.5.5.

Table 6.9. Signal data

	Raw materials					Additives		
No.	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**Strength**
1	34.27	7.10	20.08	24.30	9.48	1.17	3.60	49.77
2	26.78	21.71	15.23	23.84	7.00	1.74	3.70	53.73
3	17.01	26.04	19.65	23.16	9.41	1.12	3.60	54.10
4	23.77	22.25	15.40	25.67	7.00	2.21	3.70	54.29
7	22.11	21.71	19.91	23.84	7.00	1.74	3.70	59.14
8	20.81	21.05	19.25	26.56	7.00	1.63	3.69	59.89
9	12.18	31.64	19.91	23.84	7.00	1.74	3.70	60.59
10	19.66	23.15	21.35	22.75	7.00	2.37	3.71	61.51

6.5.4 NORMALIZATION OF SIGNAL DATA

Following the procedure discussed in Subsection 2.4.2, we normalize the individual pieces of Signal Data by subtracting the average values, shown in Table 6.8, for each of the items in the Unit Space and for each of the strengths. Normalized Signal Data is displayed in Table 6.10.

Table 6.10. Normalized signal data

	Raw materials					Additives		
No.	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**Strength**
1	12.14	−19.00	4.55	0.44	2.48	−0.57	−0.05	−6.59
2	4.65	−4.39	−0.30	−0.02	0.00	0.00	0.05	−2.63
3	−5.12	−0.06	4.12	−0.70	2.41	−0.62	−0.05	−2.26
4	1.64	−3.85	−0.13	1.81	0.00	0.47	0.05	−2.07
7	−0.02	−4.39	4.38	−0.02	0.00	0.00	0.05	2.78
8	−1.32	−5.05	3.72	2.70	0.00	−0.11	0.04	3.53
9	−9.95	5.54	4.38	−0.02	0.00	0.00	0.05	4.23
10	−2.47	−2.95	5.82	−1.11	0.00	0.63	0.06	5.15

6.5.5 COMPUTATION OF PROPORTIONAL COEFFICIENT β AND SN RATIO η

Following the procedure discussed in Chapter 2, Subsection 2.4.2, we compute proportional coefficient β and SN ratio η (duplicate ratio) item by item. The computation results are displayed in Table 6.11. If the SN ratio turns out to be negative, the value will be treated as zero in accordance with the definition given in Subsection 2.4.2 (equation 2.22).

Table 6.11. Proportional coefficient β and SN ratio η (duplicate ratio) item by item

	Raw materials					**Additives**	
β, η	**1**	**2**	**3**	**4**	**5**	**1**	**2**
β	−1.155	0.990	0.286	−0.010	−0.175	0.057	0.008
η	0.059	0.011	0.000	0.000	0.018	0.016	0.030

6.5.6 COMPUTATION OF INTEGRATED ESTIMATED STRENGTH VALUE $\hat{M}$ FOR EACH DATA ITEM

Using proportional coefficient β and SN ratio η (duplicate ratio) separately for each item, we find the integrated estimate value of the strength of each instance of the Signal Data. Integrated estimate strength value $\hat{M}_i$ for the i-th piece of Signal Data is found with the use of the equation (6.1). Note that, here, $X_{i1}, X_{i2}, \ldots, X_{i7}$ are the values of normalized raw materials 1, 2, 3, 4, and 5, and additives 1 and 2 in the i-th piece of Signal Data.

$$\hat{M}_i = \frac{\eta_1 \times \frac{X_{i1}}{\beta_1} + \eta_2 \times \frac{X_{i2}}{\beta_2} + \eta_3 \times \frac{X_{i3}}{\beta_3} + \eta_4 \times \frac{X_{i4}}{\beta_4} + \eta_5 \times \frac{X_{i5}}{\beta_5} + \eta_6 \times \frac{X_{i6}}{\beta_6} + \eta_7 \times \frac{X_{i7}}{\beta_7}}{\eta_1 + \eta_2 + \eta_3 + \eta_4 + \eta_5 + \eta_6 + \eta_7}$$

$$= \frac{0.059 \times \frac{X_{i1}}{(-1.155)} + 0.011 \times \frac{X_{i2}}{0.990} + 0 + 0 + 0.018 \times \frac{X_{i5}}{(-0.175)} + 0.016 \times \frac{X_{i6}}{0.057} + 0.030 \times \frac{X_{i7}}{0.008}}{0.059 + 0.011 + 0 + 0 + 0.018 + 0.016 + 0.030} \quad (6.1)$$

From the computation equation for integrated estimate value $\hat{M}_i$, we see that the greater the SN ratio η of an item, the more substantial the degree of contribution of its strength to the integrated estimation. Note in this connection that the SN ratios for raw materials 3 and 4 are substituted with zero so that, as it turns out, they will not be used for integrated estimation purposes.

For the first piece of data in the normalized signal shown in Table 6.10, a computation is demonstrated below of integrated estimate value $\hat{M}_1$:

$$\hat{M}_1 = \frac{0.059 \times \frac{12.15}{(-1.155)} + 0.011 \times \frac{(-19.00)}{0.990} + 0 + 0 + 0.018 \times \frac{2.48}{(-0.175)} + 0.016 \times \frac{(-0.56)}{0.057} + 0.030 \times \frac{(-0.05)}{0.008}}{0.059 + 0.011 + 0 + 0 + 0.018 + 0.016 + 0.030}$$

$$= \frac{-0.622 - 0.214 + 0 + 0 - 0.260 - 0.160 - 0.182}{0.134} = -10.706 \approx -10.71 \tag{6.2}$$

In the same manner, integrated estimate strength value $\hat{M}$ is found for each item of Signal Data; the true value (measured value) of strength M, and integrated estimate value $\hat{M}$ are shown in Table 6.12.

Table 6.12. Measured value M and integrated estimate value $\hat{M}$ for strength of signal data

Data No.	Measured Value M	Integrated Estimate Value $\hat{M}$
1	−6.59	−10.71
2	−2.63	−0.80
3	−2.26	−2.58
4	−2.07	1.44
7	2.78	0.99
8	3.53	1.09
9	4.23	5.61
10	5.15	3.59

6.5.7 COMPUTATION OF INTEGRATED ESTIMATE SN RATIO (db)

Based on the true (measured) value M and integrated estimate value $\hat{M}$ of strength in Signal Data in Table 6.12, the procedure for computing the integrated estimate SN ratio η (db) is found as shown below. The integrated estimate SN ratio η (db), which we find here, is used for the evaluation of the relative importance of items, which was discussed in Subsection 6.5.8.

Proportional equation

$$\begin{aligned} L &= M_1\hat{M}_1 + M_2\hat{M}_2 + \cdots + M_l\hat{M}_l \\ &= (-6.59)\times(-10.71) + (-2.63)\times(-0.80) + \cdots + 5.15\times 3.59 \\ &= 124.3422 \end{aligned} \tag{6.3}$$

Effective divider

$$\begin{aligned} r &= M_1^2 + M_2^2 + \cdots + M_l^2 \\ &= (-6.59)^2 + (-2.63)^2 + \cdots + 5.15^2 = 124.3422 \end{aligned} \tag{6.4}$$

Total variation

$$\begin{aligned} S_T &= \hat{M}_1^2 + \hat{M}_2^2 + \cdots + \hat{M}_l^2 \\ &= (-10.71)^2 + (-0.80)^2 + \cdots + 3.59^2 = 170.5354 \qquad (\mathrm{f}=8) \end{aligned} \tag{6.5}$$

Variation of proportional term

$$S_\beta = \frac{L^2}{r} = \frac{124.3422^2}{124.3422} = 124.3422 \qquad (\mathrm{f}=1) \tag{6.6}$$

Error variation

$$S_e = S_T - S_\beta = 170.5354 - 124.3422 = 46.1932 \qquad (\mathrm{f}=7) \tag{6.7}$$

Error variance

$$V_e = \frac{S_e}{l-1} = \frac{46.1932}{7} = 6.5990 \tag{6.8}$$

Using the above computation results, the SN ratio η for integrated estimation can be found as follows:

$$\begin{aligned} \eta &= 10\log\left(\frac{\frac{1}{r}\left(S_\beta - V_e\right)}{V_e}\right) \\ &= 10\log\left(\frac{\frac{1}{124.3422}(124.3422 - 6.5990)}{6.5990}\right) = 10\log(0.1435) = -8.43 \ (db) \end{aligned} \tag{6.9}$$

6.5.8 EVALUATION OF IMPORTANCE OF ITEMS

6.5.8.1 Layout on Orthogonal Array L_{12}

So far, we have dealt with integrated estimation using all seven items. Some of the seven items are indeed useful for the estimation, but it is thought that others are not as useful. So we proceed to evaluate the relative importance of items by assigning them to appropriate places on orthogonal array L_{12} shown in Table 6.13. For instance, in Test No. 1, because all seven items are on Level 1, these seven items are used in proceeding with the computation.

6.5.8.2 Computation of Integrated Estimate SN Ratio (db) Test Number by Number

Following the procedure discussed in Subsection 6.5.7, we find the integrated estimate SN ratio η for each test number. The results are shown in Table 6.13.

Table 6.13. Orthogonal array layout and integrated estimate SN ratios

	Raw materials					Additives						Integrated estimate
No.	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**e**	**e**	**e**	**e**	**SN ratio (db)**
1	1	1	1	1	1	1	1	1	1	1	1	−8.43
2	1	1	1	1	1	2	2	2	2	2	2	−11.24
3	1	1	2	2	2	1	1	1	2	2	2	−8.37
4	1	2	1	2	2	1	2	2	1	1	2	−9.52
5	1	2	2	1	2	2	1	2	1	2	1	−8.45
6	1	2	2	2	1	2	2	1	2	1	1	−9.82
7	2	1	2	2	1	1	2	2	1	2	1	−15.10
8	2	1	2	1	2	2	2	1	1	1	2	−19.53
9	2	1	1	2	2	2	1	2	2	1	1	−11.31
10	2	2	2	1	1	1	1	2	2	1	2	−13.96
11	2	2	1	2	1	2	1	1	1	2	2	−13.46
12	2	2	1	1	2	1	2	1	2	2	1	−17.96

6.5.8.3 Creation of a Factorial Effect Chart of Integrated Estimate SN Ratios (db)

Using the data in Table 6.13, we find an auxiliary table of (average values at each level) integrated estimate SN ratios (db); the result of the work is shown in Table 6.14. We then create, from the

Table 6.14. Auxiliary table of (averages by level to) integrated estimate SN ratios (db)

Items	Level 1	Level 2
Material 1	−9.31	−15.22
Material 2	−12.33	−12.20
Material 3	−11.99	−12.54
Material 4	−13.26	−11.27
Material 5	−12.00	−12.53
Additive 1	−12.22	−12.30
Additive 2	−10.67	−13.86

auxiliary table, a factorial effect chart, shown in Figure 6.26. From this figure, we learn that, given the significant shrinkage of the mixing ratios of raw material 1 and additive 2 from Level 1 to Level 2, the mixing ratio of raw material 1 and additive 2 plays a major part in product strength prediction.

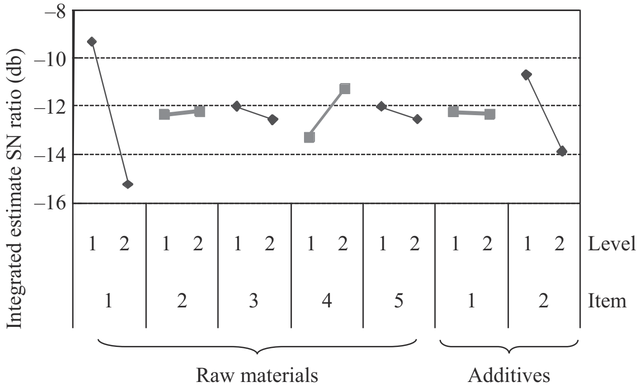


Figure 6.26. Factorial effect chart of 7 items. (Note: level 1 – item will be used; level 2 – item will *not* be used.)

6.5.8.4 Computation of Integrated Estimate Signal Data Value $\hat{M}$ under Optimum Conditions

Through a comparison of the integrated estimate SN ratios (db) in the following two cases, we will find the optimum conditions. The items (mixing ratios) to use for integrated estimation are:

Case 1: Items where the SN ratios came out positive in Table 6.11, namely raw material 1, 2, and 5 and additive 1 and 2.

Case 2: Those items where integrated estimate SN ratios came out greater on Level 1 than on Level 2 (with the factorial effect "leaning to the right") in Figure 6.26, namely raw material 1 and 5 and additive 2. (Raw material 3 will not be used because its SN ratio was zero in Table 6.11.)

Following the procedure discussed in Subsection 2.4.2, in connection with the handling of Case 2, we proceed to find the integrated estimate strength value $\hat{M}$ for data numbers 1–4 and 7–10 of the signal data. The true strength value (measured value) M and integrated estimate value $\hat{M}$ are shown in Table 6.15.

Table 6.15. Measured value M and integrated estimate value $\hat{M}$ for strength in case 2

Data No.	Measured value M	Integrated estimate value $\hat{M}$
1	−6.59	−9.93
2	−2.63	−0.53
3	−2.26	−1.61
4	−2.07	0.97
7	2.78	1.70
8	3.53	2.18
9	4.23	6.45
10	5.15	3.14

Going further, integrated estimate SN ratio (db) for Case 2 yielded –7.44 db. The two cases are compared in Table 6.16, which shows that the integrated estimate SN ratio for Case 2 turns out 0.99 db greater than in Case 1. Case 2 is thereby determined to represent the optimum condition.

Table 6.16. Comparison of the integrated estimate SN ratios (db)

Case	Used items	Integrated estimate SN ratio (db)
1	Raw materials 1,2,5 Additives 1,2	−8.43
2	Raw materials 1,5 Additives 2	−7.44

6.5.9 NTEGRATED ESTIMATION OF UNKNOWN DATA

Table 6.17 shows the strengths of two instances of newly obtained unknown data. In these cases, integrated strength estimation is performed under the optimum conditions (Case 2).

Table 6.17. Newly acquired unknown data

	Raw materials					Additives		
No.	1	2	3	4	5	1	2	Strength
<1>	23.77	22.25	15.40	25.67	7.00	2.21	3.70	Unknown
<2>	17.44	21.71	24.58	23.84	7.00	1.74	3.70	Unknown

With unknown data <1> and <2>, as in earlier examples, normalization is performed by subtracting from it the average of each item of the Unit Space. The result of this operation is displayed in Table 6.18.

Table 6.18. Normalized unknown data

	Raw materials					Additives		
No.	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**Strength**
<1>	1.64	−3.85	−0.12	1.81	0.00	0.47	0.05	Unknown
<2>	−4.69	−4.39	9.05	−0.02	0.00	0.00	0.05	Unknown

Using the computation equation discussed in Subsection 6.5.6, as in treatment of Signal Data, the integrated estimate value $\hat{M}$ of unknown data, normalized, as was shown in Table 6.17, is found as follows:

$$\hat{M}_{<1>} = \frac{0.059\times\dfrac{1.64}{(-1.155)} + 0.018\times\dfrac{0}{(-0.175)} + 0.030\times\dfrac{0.05}{0.008}}{0.059+0.018+0.030} = \frac{-0.084+0+0.188}{0.107} = 0.97 \qquad (6.10)$$

$$\hat{M}_{<2>} = \frac{0.059\times\dfrac{(-4.69)}{(-1.155)} + 0.018\times\dfrac{0}{(-0.175)} + 0.030\times\dfrac{0.05}{0.008}}{0.059+0.018+0.030} = \frac{0.240\ +0+0.182}{0.107} = 3.94 \qquad (6.11)$$

6.5.10 *COMPUTATION OF INTEGRATED ESTIMATE VALUE BEFORE NORMALIZATION*

Now that the Signal Data and unknown data have been normalized on the average value of the Unit Space, the integrated estimate value $\hat{y}$ for their pre-normalization (actual) strength is found by adding the average value M_0 (=56.36 MPa) of the Unit Space strength to the normalized integrated estimate value $\hat{M}$.

$$\hat{y} = \hat{M} + M_0 = \hat{M} + 56.36 \quad (\text{MPa}) \qquad (6.12)$$

After that, with the true strength value (actual value), y, of the unknown data known, it is now possible to display in Table 6.19 the true strength values (actual values), y, and integrated estimate values, $\hat{y}$, of the Signal Data and unknown data.

Table 6.19. Measured value y and integrated estimate value $\hat{y}$ of signal data and unknown data

Data No.		Measured value y	Integrated estimate value $\hat{y}$
Signal Data	1	49.77	46.43
	2	53.73	55.83
	3	54.10	54.75
	4	54.29	57.33
	7	59.14	58.06
	8	59.89	58.54
	9	60.59	62.81
	10	61.51	59.50
Unknown Data	<1>	56.40	57.33
	<2>	63.11	60.30

Figure 6.27 is a scatter diagram of the true strength values (measured values), y, and integrated estimate values, $\hat{y}$, of the eight pieces of Signal Data and two pieces of unknown data. The horizontal axis represents the true strength values (measured values), y, and estimated values, $\hat{y}$.

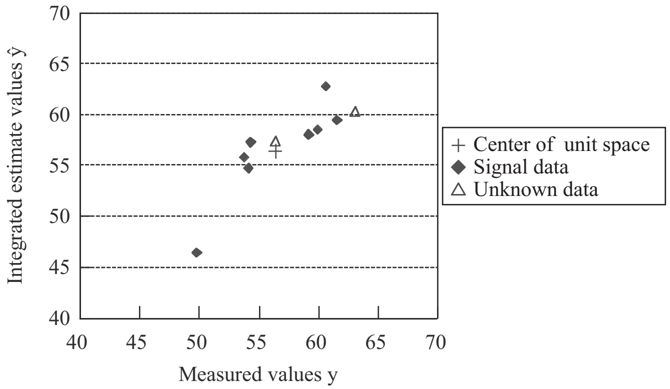


Figure 6.27. Scatter diagram of true strength values (measured values), y, and integrated estimate values, $\hat{y}$.

A look at the figure makes it clear that, given how the plotted dots are almost exactly lined up with the 45-degree straight line, both the Signal Data and unknown data are relatively closely estimated.

6.5.11 SUMMING UP THE RAW MATERIAL MIXING ISSUES

As this example illustrates, when a certain amount of data is involved, using the T Method-1 makes it possible to perform estimations with a good measure of accuracy in terms of unknown mixing ratios.

In this case, we took up tensile strength as the output and used T Method-1 as the analytical tool. The evaluation, however, must include not only tensile strength, but other output characteristics, as well, such as flexural strength and impact strength. The final mixing ratio to be adopted following the analyses mentioned above will be determined taking into consideration the results of analyses pertaining to multiple output characteristics.

Furthermore, based on the results of T Method-1 analyses, in certain cases, further test designs predicated on the use of orthogonal arrays may be elaborated upon following the acquisition of knowledge concerning the relative importance of the items.

Note 6.4 Differences between T Method-1 and the Multiple regression analysis

With respect to the data in Table 6.7, some consideration is in order as it relates to differences of the results of work with T Method-1 from the results of cases in which strength estimation was done using multiple regression analysis. Computations in the context of multiple regression analysis were performed with the help of StatWorks, a statistical operations package served by JUSE (Union of Japanese Scientists and Engineers). The correlation matrix of each item is shown in Table 6.20; the results of multiple regression analysis are given in Table 6.21.

Table 6.20. Correlation matrix

Name of items	Raw materials					Additives		Strength
	1	2	3	4	5	1	2	
Material 1	1	−0.857++	−0.176	0.19	0.333	−0.217	−0.285	−0.739+
Material 2	−0.857++	1	−0.331	−0.218	−0.493	0.303	0.154	0.588
Material 3	−0.176	−0.331	1	−0.186	0.278	−0.125	0.273	0.273
Material 4	0.19	−0.218	−0.186	1	−0.201	0.026	0.197	−0.052
Material 5	0.333	−0.493	0.278	−0.201	1	−0.784+	−0.762+	−0.663+
Additive 1	−0.217	0.303	−0.125	0.026	−0.784+	1	0.704+	0.552
Additive 2	−0.285	0.154	0.273	0.197	−0.762+	0.704+	1	0.59
Strength	−0.739+	0.588	0.273	−0.052	−0.663+	0.552	0.59	1

Table 6.21. Results of multiple regression analysis

Objective variable	Residual sum of squares	Multiple correlation coefficient	Coefficient of determination	R^{*2}
Strength	9.742	0.96	0.921	0.858
	R^{**2}	**Degree of freedom of residues**	**Residual standard deviation**	
	0.807	5	1.396	

Explanatory variable	Residual sum of squares	Variation value	Variance ratio	Partial regression coefficient
Constant term	26.121	16.379	8.406	276.359
Material 1	22.882	13.14	6.7439	−0.242
Material 2	9.187	−0.556	0.242–	–
Material 3	30.776	21.034	10.7953	0.873
Material 4	9.501	−0.242	0.1018–	
Material 5	36.567	26.825	13.7674	−4.591
Additive 1	8.738	−1.004	0.4597+	
Additive 2	18.717	8.975	4.606	−53.389

When variable selection is made consecutively with the use of multiple regression analysis, the resulting order of selection is: raw material 1, raw material 5, raw material 3, additive 2. As a result, the following multiple regression equation is found:

$$\begin{aligned} y = 276.359 - 0.242 \times \text{Material } 1 + 0.873 \times \text{Material } 3 \\ -4.591 \times \text{Material } 5 - 53.389 \times \text{Additive } 2 \end{aligned} \tag{6.13}$$

A comparison of the simple regression coefficient for strength in the correlation matrix in Table 6.20 against the partial regression coefficient in the result of the multiple regression analysis in Table 6.21 shows that raw materials 1, 5, and 3 share the same sign, whereas additive 2 carries a reversed sign. This is thought to be due to the fact that the simple correlation coefficients with respect to additive 2 and raw material 5 are large at –0.762. When T Method-1 is used, on the other hand, positive-negative sign reversal does not occur, making it easy to make sense of the equation. Also, the simple correlation coefficient of raw material 3 with respect to strength is small, but its correlation coefficients with respect to other items are also small; in view of these facts, it has been selected as an explanatory variable for the multiple regression formula. On the other hand, in the framework of T Method-1, raw material 3 is not be used for estimation if its simple correlation coefficient is small in terms of strength.

With the newly acquired unknown data shown in Table 6.15 we used both multiple regression analysis as well as the T Method-1 to perform estimations. The result is shown in Table 6.22.

Unknown data <1> is interpolated data, while unknown <2> is extrapolated data, which is located just outside of the data range used for the estimation formula. The results of the present analysis runs show that estimations performed by T Method-1 are more significantly consistent with unknown data <1>, while estimations performed using multiple regression analysis are more in line with unknown data <2>.

Because different results will be obtained depending upon the properties of the data as well as Unit Space and Signal Data selection, no conclusive determination can be made on the basis of these limited results as to which of these of the two methods used is more suitable for analysis. It may be that more depends on the properties of data. Further findings from future research will be invaluable.

Table 6.22. Comparison between estimated value obtained using multiple regression analysis and estimated value obtained using T Method-1

No.	Measured value	Estimated value $\hat{y}$ using multiple regression analysis	Estimated value $\hat{M}$ using T Method-1
<1>	56.40	54.38	57.37
<2>	63.11	63.95	60.34

Note 6.5 Why exclude unit space data from signal data?

In the context of T Method-1, Signal Data has been defined as what remains after Unit Space data is eliminated from the entire body of data (cf. Subsection 6.5.3, Table 6.9).

For information on what order of discrimination potential is in store, it is important to conduct a check using Signal Data not belonging to a Unit Space.[1] When T Method-1 is used, the location where the results congregate in a highly concentrated manner in the vicinity of average values is selected as the Unit Space. Thus, the individual data clusters in a Unit Space will be positioned in the vicinity of the center of the Unit Space. Any effort to improve the integrated estimate SN ratio, if only to a permitted maximum degree, by including Unit Space data in the Signal Data, will be an exercise in futility.

This seems to have some affinity to the concept of the "jackknife method," which, instead of seeking to elevate the multiple correlation coefficient through multiple regression analysis, computes the multiple regression equation after eliminating members from the body of specimens one by one and in order and then evaluates the residues of data not used for the computation of the multiple regression equation.

6.6. REAL ESTATE PRICE PREDICTION BY T METHOD-1[11]

The example discussed in this section is a trail-blazing case of applying T Method-1 to real estate price prediction. Mr. Sohei Yoshino and others at Yoshino Real Estate Appraisers resolved the issue of "to what extent green assets impact real estate rental fees" with the help of T Method-1. The case study focused on the Tama District of Tokyo. This is a district that has undergone a transformation from an earlier agricultural economy to an urbanized community, surviving a variety of changes in the modes of transportation, natural environment, and other life-affecting areas.

6.6.1 OVERVIEW OF REAL ESTATE PRICE PREDICTION AND ITS PURPOSE

Some of the factors that impact real estate rental fees are shown in Figure 6.28. Our purpose in conducting this analysis lies in the quantitative evaluation of the impact that the calming view of greenery such as turf-blanketed squares, tree-lined promenades, etc., the so-called green assets, on the premises of apartment complexes may have on the rental fees of real estate (monthly apartment unit rentals per square meter of living space).

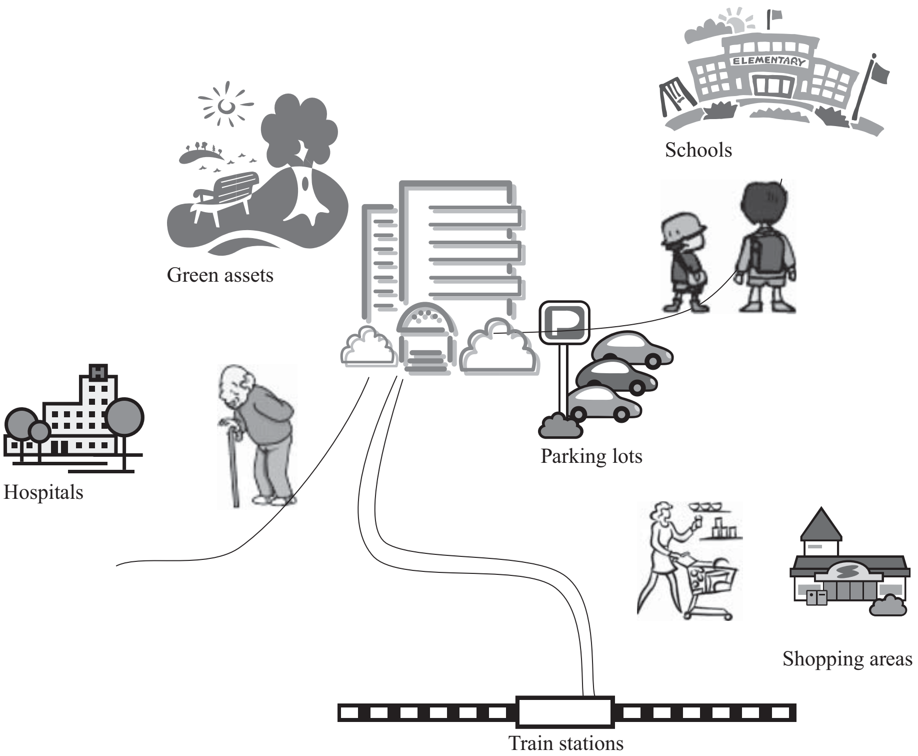


Figure 6.28. Some factors that impact real estate pricing.

6.6.2 DEFINITION OF UNIT SPACE

In real estate appraisal, a theoretical standard property called "reference unit of living space" is defined. The example in this case involves a grouping of 41 apartment complexes situated in the Tama area of Tokyo. From each of the 41 complexes, two properties were selected as reference units of living space. Accordingly, $41 \times 2 = 82$ properties were selected as targets of analysis.

Out of the entire collection of 82 properties, the one that came closest to the average of the monthly rent (in yen) per m^2 of space was selected to be the output value. Since the floor space of residences varies, the rent for the unit floor space will be used as the output.

Eighty-one items were adopted that were thought to influence the rental fees. As Table 6.23 shows, these items can be broadly broken down into 12 categories. The category "Access Conditions," for example, includes time necessary to access the nearest train station, shopping areas, schools, hospitals, etc.

It should be noted, however, that there were nine items that showed the same value for each of the 82 properties. Those items, therefore, were excluded from the analysis, 72 items remained at the end ($81 - 9 = 72$). The values of these items reflect five levels of score assignment and further specialized evaluation.

Table 6.23. Major classification items adopted for rent prediction

Condition	No.	Condition	No.	Condition	No.
Access	1–16	Bldg configuration, structure	46–52	Scenic considerations	74–75
Road layout	17–21	Bldg floor plan	53–66	Functionality for common usability	76–77
Neighborhood settings	22–30	Handicap accessibility	67–71	Urban services	78–79
Design of grounds	31–45	Micro-climate control	72–73	Nature	80–81

Note: The numbers (No.) are given as subcategory indices.

6.6.3 DEFINITION OF SIGNAL DATA AND NORMALIZATION

With one property assigned to the Unit Space that was subtracted from the entire body of 82, the properties that make up the Signal Data now number 81. Following the procedure shown in Subsection 2.4.2, we performed normalization for items 1–72, one by one, by subtracting "(1) value of Unit Space" from the "(2) value of Property 1." We extracted, as a sample, property 1 from the Signal Data made up of the 81 properties, with the normalization results shown in Table 6.24.

Table 6.24. Normalization of signal data (property 1 taken as an example)

Item No.	Item	(1) Value of unit space	(2) Value of property 1	After normalization (2)–(1)
1	Access time to nearest station	5	3	−2
2	Access time from nearest station to terminal	2	3	1
3	Shopping areas	4	4	0
4	Schools	4	5	1
5	Hospitals	4	4	0
⋮	⋮	⋮	⋮	⋮
69	Evacuation route; community evacuation center	2	3	1
70	Green buffer zone: y/n	5	5	0
71	Greenery-water continuum	3	3	0
72	Diversity of green cover	3	5	2

6.6.4 COMPUTATION OF PROPORTIONAL COEFFICIENT β AND SN RATIO η OF SIGNAL DATA

With the computation formula shown in Subsection 2.4.2, using the Signal Data normalized as just discussed, we found the proportional coefficient, β, and the SN ratio, η (duplicate ratio), item by item. Table 6.25 shows partial results of the computations. Where the SN ratio η was negative, it was substituted with a zero to conform to the definition of expression (2.22) in Subsection 2.4.2.

6.6.5 COMPUTATION OF INTEGRATED ESTIMATION OF SIGNAL DATA

From the proportional coefficient β and SN ratio η of each item, we computed the integrated estimate rental fee for each of the 81 properties, using the integrated estimation formula discussed in Subsection 2.4.2, which is treated as Signal Data. Subsequently, we evaluated to what degree the estimated and actual values of the rental fees matched. More specifically, as Figure 6.29 shows, we used a scatter diagram to illustrate how the true value (actual value) of the Signal Data,

Table 6.25. Proportional coefficient β and SN ratio η (duplicate ratio) of each item

Item No.	Item	S_β	S_T	V_e	Proportional coefficient β	SN ratio η (duplicate ratio)
1	Access time to nearest train station	0.62	250.00	3.08	3.52E-06	0
2	Access time from nearest station to terminal	2.02	86.00	1.04	6.38E-06	1.91E-11
3	Shopping areas	7.16	154.00	1.81	–1.20E-05	5.93E-11
4	Schools	1.51	40.00	0.48	5.50E-06	4.37E-11
5	Hospitals	0.37	134.00	1.65	2.71E-06	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
69	Evacuation route; communal evacuation center	0.24	122.00	1.50	2.19E-06	0
70	Green buffer zone: y/n	4.00	664.00	8.15	8.97E-06	0
71	Continuity of greenery and water	0.54	154.00	1.89	3.31E-06	0
72	Green cover diversity	9.06	194.00	2.28	1.35E-05	5.96E-11

represented by the horizontal axis, and the integrated estimate value, represented by the vertical axis, stand in relation to each other. If the estimated value and the actual value are in complete agreement, then their reciprocal relationship is expressed by a straight line with a gradient of 1. From this scatter diagram, we can appreciate that the estimation hits the mark rather closely, even though the gradient falls somewhat short of 1 and thus a completely linear relationship has not been achieved.

We computed the integrated estimate SN ratio following the procedure discussed in Subsection 2.4.2. By making appropriate use of the SN ratio, it is possible to quantitatively evaluate to what extent the actual rent value has been accurately estimated. In our study, we achieved an SN ratio of 20.61 (db). This value provides useful hints when compared with other computation methods for estimation accuracy. Note in passing that the SN ratio of the integrated estimation yielded by the TS Method, one of several other computation methods that have been proposed in the context of the MT System, is 13.21 (db), and that it has been confirmed that T Method-1 exhibits superior performance in estimation accuracy.[9]

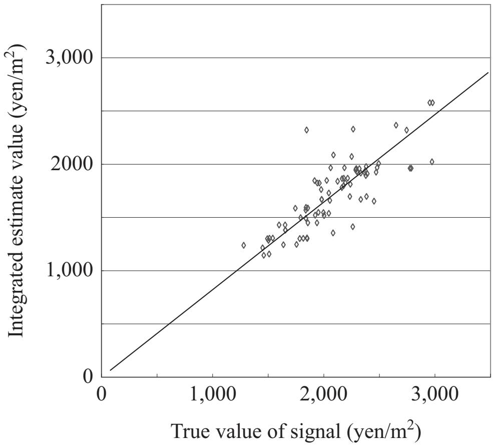


Figure 6.29. Scatter diagram of true values (actual values) and integrated estimate values of rental fees.

6.6.6 ANALYSIS OF THE IMPACT GREEN ASSETS EXERT ON RENTAL FEES

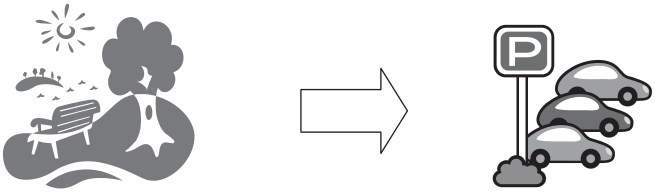
In order to verify how green assets impact the rental fees of real estate properties, we selected one property from the entire group of properties and forecast the variation in rent as a function of green pattern modification.

The following features characterize the apartment complex grounds where the property selected stands:

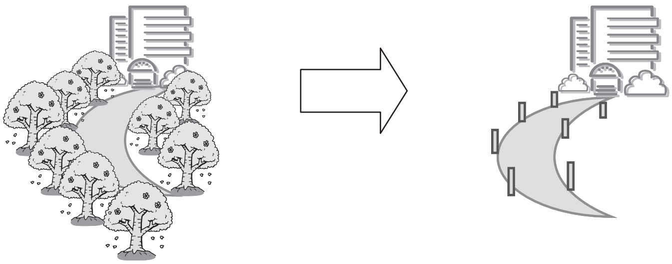
(a) a circular grassy plaza in the middle of the apartment complex;
(b) an esplanade dotted with cherry trees forming a loop along the premises of the complex.

Thanks to these two value-added features, this apartment complex has garnered a respectable score, ranking among the greenest properties in the area. At this point, as shown in Figure 6.30, we varied points (a) and (b) and performed trial calculations to see how these factors might impact the rental fees of the property.

We can find out "how much the rent will go down if the quality of the green assets deteriorate from the current standard" by changing the pattern of greenery. Since some degree of latitude is to be allowed in the variations of (a) and (b), a lower limit as well as an upper limit were placed on the decline in value brought about by a qualitative change, as shown in Table 6.26, and the outcome was analyzed using T Method-1. The "lower limit" and "upper limit" represent the following thinking:



(a) The circular grassy plaza in the middle of the premises turned into a parking lot



(b) The esplanade dotted with cherry trees on the premises turned into a treeless trail

Figure 6.30. Concept of green pattern transformation.

Table 6.26. Upper and lower limits of green assets transformation

Item No.	Item	Lower limit	Upper limit
63	Green cover ratio	4	4
64	Abundance of green shade	4	5
65	Scenic value	3	4
66	Scenic streets	3	4
67	Continuity with surrounding paths	3	3
68	Community interaction; flower viewing areas	3	3
69	Evacuation routes; communal evacuation center	3	3
70	Presence/absences of a green buffer zone	4	4
71	Continuity of greenery and water	3	3
72	Green cover diversity	4	5

Lower limit: Appraisal based on the assumption of the impact of the scenic value of greenery to be at its lowest.

Upper limit: Appraisal based on the assumption of the impact of the scenic value of greenery to be at its highest.

The findings from this analysis are summarized in Table 6.26 as indices with the current rent level set at 100. This figure makes it clear that, if green assets diminish, the rent may fall 0.5% to 1.7% as shown in Figure 6.31.

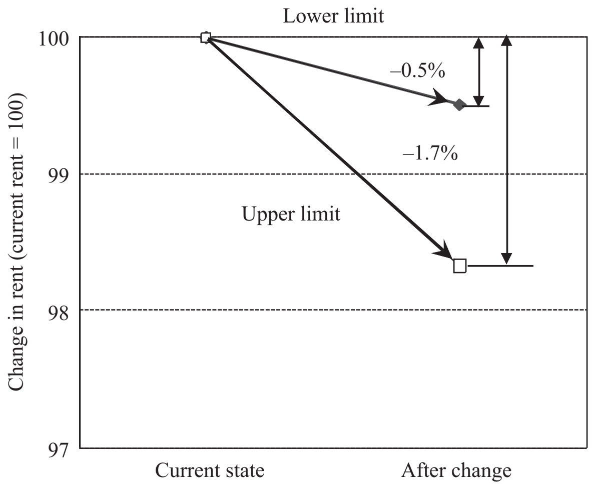


Figure 6.31. Change in rent reflecting a decrease in green assets.

6.6.7 MERITS OF T METHOD-1 AND PENDING ISSUES

Real estate evaluation and appraisal is often beleaguered by the high cost of data collection and the difficulty of obtaining transaction data. Data can be scarce, in some cases, making it even impossible to perform a multiple regression analysis. This is where T Method-1 comes in handy with its ability to provide highly accurate analyses even with limited data, offering great operational advantages.

For the purposes of this analysis, we selected from among the 82 properties the one property with rent closest to the average of all monthly rents and set it as the Unit Space. However, there is the possibility we can enhance the accuracy of the estimation by computing an average of multiple pieces of data plotted in the vicinity of the average value and setting that up as the Unit Space. The reason for this is that the reliability of the Unit Space will augment when the computation depends not on the rent of just the one property closest to the average monthly rent, but on the average of plural units of data found in the vicinity of the average monthly rent.

Note, however, that, even if a green asset contributes to increased real estate rental fees, the real estate may still end up assuming a negative economic value if the green assets management cost outweighs the benefit of the green contribution. Evaluation of the economic value of a real property that includes such considerations remains an issue to be addressed in the future.

APPENDIX A

DIFFERENCES BETWEEN THE MT SYSTEM AND ARTIFICIAL INTELLIGENCE

The following are the three most commonly used methods in the pattern recognition process:

1. matching-based method;
2. method based on artificial intelligence theory application; and
3. method based on the application of statistical mathematics.

The MT System may be regarded as belonging to method (3) here. Within the framework of the MT System, several computation methods have been proposed. One of them, the MT Method, utilizes a statistical theory called "Mahalanobis Distance." We will briefly discuss the difference between the MT Method and artificial intelligence below, to be followed by a study of how the MT Method differs from traditional statistical mathematics.

A.1 THE LEARNING PROCESS IN THE CASE OF ARTIFICIAL INTELLIGENCE

A representative example of pattern recognition based on artificial intelligence theory is an Artificial Neural Network (ANN). An ANN is made to learn the model data set, or "teaching data," repeatedly. This resembles the process of a child's learning things little by little through repetition. As a result of the learning, a recognition function is formed. The recognition function is a function for the conversion of input data into a recognition result. When recognition target data, that is, unknown data, are fed into the ANN which has assimilated the learning, the ANN outputs the recognition result.

A simplified schematic of an artificial neural network is shown in Figure A.1. Each circle (○) in the diagram represents a neuron, with the synapses indicating the inter-neurally connected relationships, shown with lines. The state of the neurons and the thickness of the synapses undergo a gradual change as the learning is repeated.

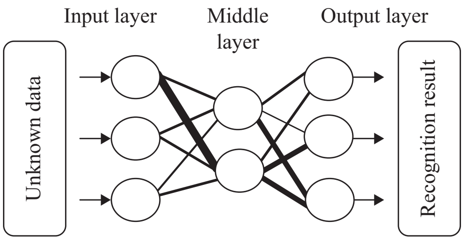


Figure A.1. Neural network (the state of neurons and synapses changes with learning).

A.2 PARAMETERS OF ARTIFICIAL INTELLIGENCE AND THE PROPERTIES OF RECOGNITION RESULTS

There are a good many parameter settings, such as the network configuration for the initial state, number of neurons, and many others, when an ANN is to be set. These parameters are set by the ANN user according to the user's needs, and so a "learned" ANN, it can be said, generally assumes an individuality profile of its own (Figure A.2). In other words, each ANN differs from others in learning speed and recognition flexibility. This feature seems to remind us of the ANN's resemblance to man.

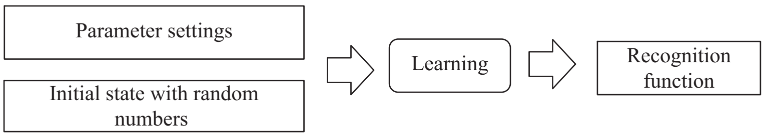


Figure A.2. Artificial intelligence learning process.

A.3 DIFFERENCES IN PROPERTIES BETWEEN THE MT METHOD AND ARTIFICIAL INTELLIGENCE

An ANN is equipped with the unique feature of being able to perform recognition flexibly, even when it has been fed limited teaching data. For instance, if it has learned only one set of numerical data, say 0–9, it can learn patterns and produce relatively accurate recognition results. From a conceptual standpoint, it can be said that, with the use of given pedagogic data, a recognition function with a certain latitude in the contour is formed. With the MT Method, however, it is not possible with one set of data to form a recognition function. In other words, it is necessary to have a mathematically reliable quantity of data in place.

On the other hand, in terms of the meaning of the recognition function, the MT Method brings clarity to the input-output (cause-and-effect) accountability. With artificial intelligence the result of learning becomes sealed in a black box. The post-learning state of neurons and synapses is a result obtained based on the rules of learning, making it difficult to explain in a logical fashion why a given result appears the way it does. But, with the MT Method, it is possible to explain the meaning of the recognition function in mathematical terms.

Recognition function formation is generally performed at a faster speed with the MT Method than with artificial intelligence. While artificial intelligence needs repeated processing of computations, the MT Method, relatively speaking, requires fewer computational operations.

Let it be noted, in passing, that with the MT Method, the following computation speeds have been recorded on a high-speed personal computer:

Given number of features: 1,000; number of samples: 2,000
Recognition function formation time: 15 sec
Recognition time: 0.01 sec/sample

Table A.1 summarizes the salient differences between the MT Method and artificial intelligence. There are a number of requirements that the pattern recognition computing system is expected to satisfy, such as reliability of results, ease of operation, flexibility, computation speed, and data add ability. The optimum tool should be chosen that best suits the given purposes after all these relevant points and factors have been taken into consideration.

Table A.1. Salient differences between the MT Method and artificial intelligence

Item	MT Method	Artificial intelligence
Principle recognition function formation tenet	Computation based on correlation matrices	Pedagogic data learned based on learning convergence theory
Amount of data needed for recognition function formation	Theoretically required amount of data exists	Even one piece of data (a single set) will suffice
Nature of recognition function	Explicit and uniquely interpretable	Not explicit and individualistic
Computation time	Short	Long

APPENDIX B

DIFFERENCE BETWEEN THE MT SYSTEM AND TRADITIONAL STATISTICAL THEORY

The MT Method, one of several tools of computation that constitute the MT System, makes applied use of Mahalanobis Distance, which is utilized in the multivariate control chart as well as in the discriminate function. The T Method, at first glance, appears to resemble regression analysis. In this section, we will discuss how the MT Method and the T Method differ from traditional statistical theory.

B.1 DIFFERENCE BETWEEN THE MT METHOD AND THE MULTIVARIATE CONTROL CHART

B.1.1 OVERVIEW OF THE CONTROL CHART AND THE MULTIVARIATE CONTROL CHART

Before studying the similarities and differences between the MT Method and the multivariate control chart, let's first take a look at the commonly used control chart (Figure B.1). A control chart sets a pair of statistically determined control limit lines used to evaluate the condition of a process in progress. The process is regarded as being "within the bounds of stable progression" if the statistic (an average, etc.) derived from an observed value is within the control limit lines and is deviation-free. If more than one statistic is being followed, one control chart will be created for each statistic.

A multivariate control chart is used to aggregate a plurality of observed values into one single value and control it. A *standard state* is defined on the basis of observed values for the normal state, and then the distance from that state to the object of control is measured. Based on the distance found, it becomes possible to discover any abnormality due to any deviation from the existing

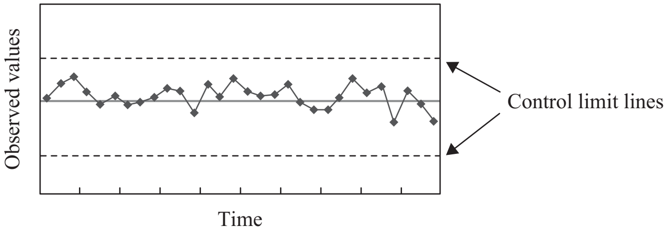


Figure B.1. Example of a control chart.

correlation. For instance, let's look at a case in which, as shown in Figure B.2, control is being performed using two observed values. If we define the control limit with dotted lines for the two observed values and perform control on each of the observed values, point A will be regarded as being within normal bounds. But if values observed in normal time are scattered within the elliptical circle, then point A cannot be said to be normal. In such a case, we can tell that point A is not normal, by taking the correlation between observed values into consideration. In a multivariate control chart, we use the Mahalanobis Distance because we aggregate multiple observed values into a single value to perform an evaluation that takes the correlation into consideration.

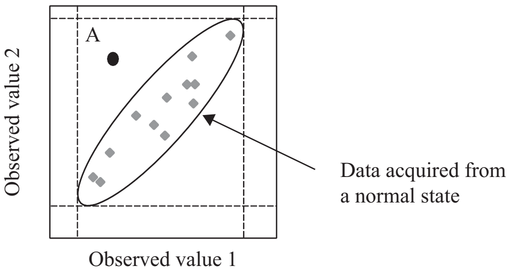


Figure B.2. Control using a bivariate control chart.

B.1.2 DIFFERENCES BETWEEN THE MT METHOD AND THE MULTIVARIATE CONTROL CHART

The MT Method defines the "Unit Space" and measures the distance from its center to the object data as a Mahalanobis Distance. It has been said that the difference between the MT Method and the multivariate control chart is difficult to understand because of certain mathematical features common to both. It is nonetheless true that the MT Method and the multivariate control chart significantly differ in terms of purpose and technical organization. The main areas of difference between them are examined below.

B.1.2.1 Difference in Purpose

The multivariate control chart represents a system designed to compensate for the shortcomings of the generic control chart. On the other hand, the MT Method is a system originally designed to serve the purpose of pattern recognition. Thus, the MT Method specifically addresses areas such as medical treatment, characteristics inspection, image inspection, estimation of unknown circumstances, prediction of future situations or developments, etc. These are items that are not addressed by the multivariate control chart.

B.1.2.2 Differences in Concept and Organizational Technique

- **Difference in Definition of the Normal State**: With the MT Method, the Unit Space is defined using *homogeneity with respect to the objective* as the standard so as to secure stable recognition. Homogeneity may be paraphrased as a "normal state," "state of constancy," "state of high density," etc. In actual practice, a normal state is used as the Unit Space in many cases, but the MT Method is based on the idea that the homogeneity factor is what the reliability of recognition depends on most importantly. The multivariate control chart, on the other hand, looks to the normal state as the standard and does not find any meaning in the presence or absence of the homogeneity factor.
- **Proposal of Feature Extraction Technique**: With pattern recognition, how successfully feature extraction is carried out determines the result of the whole process. When Dr. Taguchi introduced the MT Method in the Quality Engineering journal in 1995, he also introduced another feature extraction method of his own, which integrated feature extraction and recognition processing. He called it MTS (which eventually evolved into the present-day MT Method). The multivariate control chart, on the other hand, incorporates no concept such as feature extraction and utilizes observed data in an almost raw form.
- **Proposal of an Evaluation Criterion for the Recognition System**: The MT System includes a success/failure quantification proposal for the recognition system. Use is made here of the concept of SN ratio (signal-to-noise ratio), which provides a functionality evaluation yardstick in Quality Engineering or Taguchi Methods. In pattern recognition, generally speaking, the recognition result is influenced by various factors, ranging from measuring tool selection to recognition software parameter settings. Use of the SN ratio makes it possible to evaluate the appropriateness of the entire recognition system.
- **Proposal of Item Selection**: Item selection refers to the procedure of selecting a variable that is effective for the purposes of abnormality detection. It will help eliminate unnecessary variables so as to heighten the abnormality detection sensitivity. No proposals of this kind are found in connection with the multivariate control chart.
- **Proposal of a Diagnostic Method for Causes of Abnormality**: With the use of the item selection procedure, the causes of an episode of abnormality can be diagnosed when one occurs. When target data (unknown data) is determined to be abnormal, it will thus be possible to deduce which variable is causing the abnormality. This proposal has the effect of heightening the efficiency of the recognition system operation.

B.1.2.3 Difference in Dealing with Multicollinearity

When the absolute value of a correlative coefficient between different variables is "1," or when a variable is the sum of other variables, there arises the issue of *multicollinearity*. If multicollinearity is present, it becomes impossible, in the process of computing Mahalanobis Distance, to execute the computation of inverse matrices of correlation matrices. From the mathematical standpoint, it has been recommended to discard one or the other of the variables in such a case. Practically speaking, though, it may lose the ability to detect abnormalities, so some type of remedy is necessary.

This point can be explained in a bit more in detail using a simple example. Imagine two flow sensors, A and B, placed in a flow channel, as in Figure B.3, which show the same observed value at a time of "normalcy." That is to say, A and B have a linear correlation and pass the same start point. Their correlative coefficient is 1, and there exists a multicollinearity between them. But if a leak occurs in the flow at some point between A and B, the observed value is plotted at a point outside of the straight line. In this situation, if one of the sensors is removed or disabled, the abnormality will not be detected before a large amount of flow has leaked. The same will apply to a situation where, as shown in Figure B.4, the flow channel merges with another flow passage. Multicollinearity is precisely where relevant information is to be found. As Dr. Taguchi himself once stated "significant information lies in multicollinearity." It sometimes occurs that multicollinearity prevails at a time of normalcy but the fact that it then collapses when disturbed by an abnormality is seen occurring in actual operational settings.

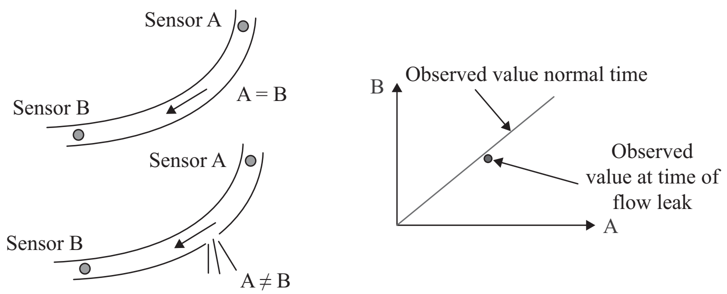


Figure B.3. Example where multicollinearity assumes importance (linear correlation in normal time collapsing when abnormality occurs.

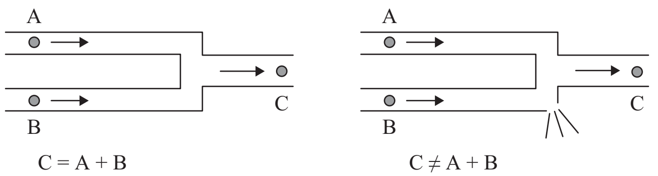


Figure B.4. Another instance of multicollinearity.

With the MT System, a proposal has been made for a proactive approach to multicollinearity. This stance distinguishes it from statistical mathematics including the multivariate control chart and multiple regression analysis.

The points discussed above are summed up in Table B.1, which will clarify how the MT Method differs from the multivariate control chart and why it constitutes a category of its own in terms of theory and technology.

Table B.1. Differences between the MT Method and the multivariate control chart

Item	MT Method	Multivariate control chart
Purpose (areas of application)	Pattern recognition, prediction	Process control
Definition of normal state	Homogeneous state	Normal state
Result evaluation	SN ratio used	–
Item selection	Yes	–
Diagnostic for causes of abnormality	Yes	–
Feature extraction	Yes	–
How multicollinearity is addressed	Proactively addressed	Avoided

B.2 DIFFERENCE BETWEEN DISCRIMINATE ANALYSIS AND THE MT METHOD

Mahalanobis Distance has traditionally been used extensively in *discriminant analysis*. Discriminant analysis is a tool for determining to which existing group a given individual belongs. For instance, if we have two groups, A and B, as in Figure B.5, we can use Mahalanobis Distance to determine to which group the object C belongs.

With the MT Method, as in Figure B.6, there is only one group to define, and the distance between the object (target) data and the center of the group is computed; the MT Method is different from discriminant analysis in that regard.

Moreover, distance “1” is the unit quantity in the MT System. This is because the average value of the Mahalanobis Distance in the Unit Space turns out to be “1.” Figure B.7 is a conceptual representation which, based on

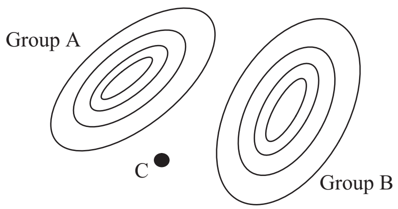


Figure B.5. Concept of discriminant analysis.

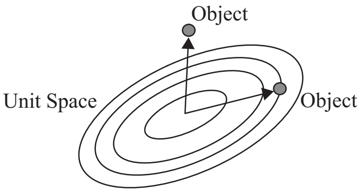


Figure B.6. Concept of group and distance using the MT Method.

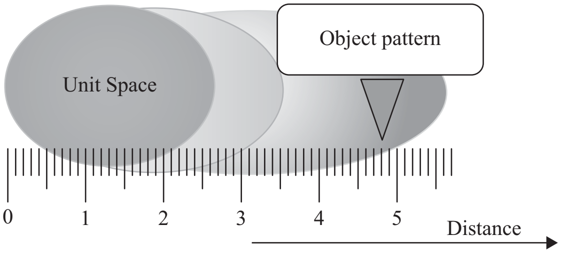


Figure B.7. Definition of unit quantity and distance scale (The MT Method treats pattern recognition as "measurement").

the Unit Space, will ultimately define the scale (yardstick) for pattern distance measurement. Conceptualization along this line of reasoning is unique to the MT Method.

B.3 DIFFERENCES BETWEEN T METHOD-1 AND MULTIPLE REGRESSION ANALYSIS

T Method-1 is a theory for prediction and estimation of an output value (objective variable) based on a multivariate and serves the same purpose as multiple regression analysis. The two, however, are considerably different in computational process. Here we will delve into some particulars as to what distinguishes T Method-1 from multiple regression analysis.

B.3.1 CONCEPT BEHIND MULTIPLE REGRESSION ANALYSIS AND COMPUTATION FORMULA

If, as shown in Table B.2, we express the multiple regression model with the objective variable represented as y and the explanatory variable as $(x_1, x_2, \ldots, x_k)$, we will get:

$$y = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_k x_k \tag{B.1}$$

If we express it as an orthogonal polynomial equation, we will have:

$$y = \overline{y}' + b_1(x_1 - \overline{x}'_1) + b_2(x_2 - \overline{x}'_2) + \cdots + b_k(x_k - \overline{x}'_k) \tag{B.2}$$

As is clear from equation (B.2), a plain defined by a multiple regression equation always passes the center of all observed points (the item-by-item average of all data in (a) of Table B.2), i.e., $(\overline{x}'_1, \overline{x}'_2, \ldots, \overline{x}'_k, \overline{y}')$.

Here, average values are marked with a prime (′), used to denote the average value of each of the items under "(a) All data" in Table B.2 and distinguish it from the average value of each of the items in "(b) Data samples for Unit Data."

Table B.2. Relationship between multiple regression analysis data and T Method-1 data

Data for multiple regression analysis

(a) All data samples

No.	Item 1	Item 2	…	Item k	y
1	X_{11}	X_{12}	…	X_{1k}	y_1
2	X_{21}	X_{22}	…	X_{2k}	y_2
…	…	…	…	…	…
m	X_{m1}	X_{m2}	…	$X_{m\,k}$	y_m
m+1	$X_{m+1\,1}$	$X_{m+1\,2}$	…	$X_{m+1\,k}$	y_{m+1}
m+2	$X_{m+2\,1}$	$X_{m+2\,2}$	…	$X_{m+2\,k}$	y_{m+2}
…	…	…	…	…	…
m+n	$X_{m+n\,1}$	$X_{m+n\,2}$	…	$X_{m+n\,k}$	y_{m+n}
m+n+1	$X_{m+n+1\,1}$	$X_{m+n+1\,2}$	…	$X_{m+n+1\,k}$	y_{m+n+1}
…	…	…	…	…	…
N–1	$X_{N-1\,1}$	$X_{N-1\,2}$	…	$X_{N-1\,k}$	y_{N-1}
N	$X_{N\,1}$	$X_{N\,2}$	…	$X_{N\,k}$	y_N
Average	X'_1	X'_2	…	X'_k	y'

Note: Plots of output y already reordered in ascending order.

Median range

Remainder

Data for T Method-1

(b) Data samples for unit data

No.	Item 1	Item 2	…	Item k	y
m+1	$X_{m+1\,1}$	$X_{m+1\,2}$	…	$X_{m+1\,k}$	y_{m+1}
m+2	$X_{m+2\,1}$	$X_{m+2\,2}$	…	$X_{m+2\,k}$	y_{m+2}
…	…	…	…	…	…
m+n	$X_{m+n\,1}$	$X_{m+n\,2}$	…	$X_{m+n\,k}$	y_{m+n}
Average	$\overline{X}_1$	$\overline{X}_2$	…	$\overline{X}_k$	$y = M_0$

(c) Signal data

No.	Item 1	Item 2	…	Item k	y
1	X_{11}	X_{12}	…	X_{1k}	y_1
2	X_{21}	X_{22}	…	X_{2k}	y_2
…	…	…	…	…	…
m	X_{m1}	X_{m2}	…	$X_{m\,k}$	y_m
m+n+1	$X_{m+n+1\,1}$	$X_{m+n+1\,2}$	…	$X_{m+n+1\,k}$	y_{m+n+1}
…	…	…	…	…	…
N–1	$X_{N-1\,1}$	$X_{N-1\,2}$	…	$X_{N-1\,k}$	y_{N-1}
N	$X_{N\,1}$	$X_{N\,2}$	…	$X_{N\,k}$	y_N

l

Note: (c) Signal Data is what remains after (b) is subtracted from (a).
Number of data items l is (N – n).

(d) Normalized signal data and the integrated estimate values

No.	Item 1	Item 2	…	Item k	M_i	M_{i1}	M_{i2}	…	M_{ik}	M_i
1	X_{11}	X_{12}	…	X_{1k}	M_1	M_{11}	M_{12}	…	M_{1k}	M_1
2	X_{21}	X_{22}	…	X_{2k}	M_2	M_{21}	M_{22}	…	M_{2k}	M_2
…	…	…	…	…	…	…	…	…	…	…
m	X_{m1}	X_{m2}	…	$X_{m\,k}$	…	M_{m1}	M_{m2}	…	$M_{m\,k}$	M_m
m+n+1	$X_{m+n+1\,1}$	$X_{m+n+1\,2}$	…	$X_{m+n+1\,k}$	M_{m+n+1}	$M_{m+n+1\,1}$	$M_{m+n+1\,2}$	…	$M_{m+n+1\,k}$	M_{m+n+1}
…	…	…	…	…	…	…	…	…	…	…
N–1	$X_{N-1\,1}$	$X_{N-1\,2}$	…	$X_{N-1\,k}$	M_{N-1}	$M_{N-1\,1}$	$M_{N-1\,2}$	…	$M_{N-1\,k}$	M_{N-1}
N	$X_{N\,1}$	$X_{N\,2}$	…	$X_{N\,k}$	MN	$M_{N\,1}$	$M_{N\,2}$	…	$M_{N\,k}$	M_N
Proportional coefficient	β_1	β_2	…	β_k						
SN ratio	η_1	η_2	…	η_k						

l

The partial regression coefficient of multiple regression analysis, $a_1, a_2, \ldots, a_k$, is computed in such a way that the sum of squares of the residual difference between the predicted value of y and the measured value will come out at its smallest. Toward this end, the computation is performed after the inverse matrix is found of the sum-of-squares-and-products matrix with inter-item correlations included. Note in working through this computation that there is a restriction that the total number of pieces of data, N, must be greater than the number of items, k. Moreover, the issue of inter-item multicollinearity may surface.[9] In other words, if the absolute value of an inter-item correlation turns out to be 1, the computation is impossible, or the accuracy of computation may become unstable. It is thus necessary to take a careful look at the scatter diagram or the values of the correlation coefficients to check for the presence, or lack of multicollinearity.

B.3.2 CONCEPT BEHIND T METHOD-1

With T Method-1, the number of all items of data, N, is first divided into n number of Unit Space samples (see (b), Table B.2) and Signal Data/(=N−n) number (see (c), Table B.2). Note here that the Unit Space samples must be selected out of a desirably homogeneous and dense congregation of output values (objective variables y) in the middle or near the middle of all data.[3]

After the Unit Space has been decided, normalization is performed by subtracting, as shown below, the averages of each item and output of each Unit Space sample (see averages under (b), Table B.2), i.e., $(\bar{x}_1, \bar{x}_2, ..., \bar{x}_k, \bar{y})$, from the Signal Data.

$$X_{ij} = x_{ij} - \bar{x}_j \qquad (i=1,\ldots,m,m+n+1,\ldots,N) \quad (j=1,2,\ldots,k) \tag{B.3}$$

$$M_i = y_i - \bar{y} \qquad (i=1,\ldots,m,m+n+1,\ldots,N) \tag{B.4}$$

Due to normalization, the general estimation formula passes the center $(\bar{x}_1, \bar{x}_2, ..., \bar{x}_k, \bar{y})$ of the Unit Space.

We will compute next the proportional coefficient β_j and SN ratio η_j (duplicate ratio), item by item. Namely:

1. With output (signal) value M_i assigned to the horizontal axis and item value X_{ij} to the vertical axis, suppose a regression equation that passes origin $X_{ij} = \beta_j M_i$.
2. Find proportional coefficient β_j, and evaluate the scatter from the proportional equation by how it relates to SN ratio η_j.
3. And find output estimated value $\hat{M}_{ij} = \dfrac{X_{ij}}{b_j}$ item by item for the signal data, and, following weighting using SN ratio η_j, find integrated estimate value M_i through addition as follows:

$$\hat{M}_i = \frac{\eta_1 \times \dfrac{X_{i1}}{\beta_1} + \eta_2 \times \dfrac{X_{i2}}{\beta_2} + \cdots + \eta_k \times \dfrac{X_{ik}}{\beta_k}}{\eta_1 + \eta_2 + \cdots + \eta_k} \qquad (i=1,\ldots,m,m+n+1,\ldots,N) \tag{B.5}$$

The preceding discussion is summarized in Table B.2 (d).

For the Signal Data, as well, an integrated estimation is performed of the pattern difference from the center of the Unit Space using the β_j and η_j of each item.

T Method-1 does not use in the computations discussed above processes calling for the use of the correlation matrix or its inverse matrix and so incurs neither restrictive conditions such as "number of Unit Space samples, n > number of items, k" nor issues of multicollinearity. What this means is that T Method-1 makes it possible, taking maximum advantage of the information given, to analyze issues with limited available data conventionally, which have heretofore been considered not to be analyzable.

T Method-1 lets the analysis object reveal its properties without the need to submit it to new experiments; instead, the method usefully takes advantage of accumulated past test data and day-to-day manufacturing data. It should be noted, however, that, to the extent that selection of Unit Space data is left to the discretion of the operator, the ultimately acquired estimated value is influenced by this. The differences between T Method-1 and multiple regression analysis are summarized in Table B.3.

Table B.3. Comparison of T Method-1 and multiple regression analysis

	T Method-1	Multiple regression analysis
Unit Space	Selection made out of homogeneous, dense population with the output value in the median range	"Unit Space" concept not adopted. Total data used to compute regression formula.
Restrictive conditions	Number of Unit Space samples n≧1. Number of Signal Data samples 1≧2. (Multicollinearity does not occur.)	Restriction in effect that total number of items n>number of items k. No solution possible if multicollinearity is present. Elimination of items may make it possible, but the influence of important items may become impossible to analyze.
Correlation between items	Correlation between items *not* made use of. Single regression is used involving Signal Data outputs and items (proportional equations with zero as the reference point). If correlation close to "1" exists, it may affect accuracy of the integrated estimate value.	Correlation between items made use of. If correlation close to "1" exists, the signs of a partial regression coefficient and a single regression coefficient will not match.
Adopted evaluation function	Integrated estimate SN ratio η	Multiple correlation coefficient, or, multiple correlation coefficient adjusted for the degrees of freedom

APPENDIX C

SUPPLEMENTARY CONSIDERATIONS CONCERNING MATHEMATICAL FORMULAS

Computation formulas used in the MT Method were discussed in Chapter 2. The standard deviation σ used here is computed with the following equation:

$$s = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n}} \tag{C.1}$$

With reference to the n used in equation (C.1), note that statistical computation in general use is $n - 1$ instead; MT Method adopts a different practice.

APPENDIX D

Strategy to Use When Data Incorporates Unmeasured Values

When distances or estimated values of target data are being measured, variables (items) that have escaped measurement occur from time to time. Let's suppose for example that, as illustrated in Table D.1 (a), samples *i* and *j* incorporate unmeasured values. Sample *i* contains unmeasured variable 3, for example. The presence of unmeasured values makes it impossible to proceed further with computation, and measures need to be taken to compensate for this deficient value.

In Table D.1 (a), we notice that sample *i* has variable 3 missing, and in sample *j*, variable 2 is likewise missing. A number of studies have been made in the quest for some appropriate method of compensation. For example, a method has reportedly been proposed that uses a Unit Space average as a means of compensation.[6] The proposal takes note of the point that unmeasured values are apt not to occur with important data and, as shown in Table D.1 (b), calls for substituting the unmeasured values with the Unit Data average values. The issues also depend upon the nature of the problem and the variable in question, and there continues to be room for further examination.

Table D.1. Compensating for unmeasured data

Variable 1	Variable 2	Variable 3	…	Variable k
(a) *Unmeasured values*				
x_{i1}	x_{i2}	unmeasured	…	x_{ik}
x_{j1}	unmeasured	x_{j3}	…	x_{jk}
$\overline{x}_1$	$\overline{x}_2$	$\overline{x}_3$	…	$\overline{x}_k$
(b) *Examples of compensation*				
x_{i1}	x_{i2}	$\overline{x}_3$	…	x_{ik}
x_{j1}	$\overline{x}_2$	x_{j3}	…	x_{jk}

APPENDIX E

FUSION WITH ARTIFICIAL INTELLIGENCE AND OTHER RESOURCES

In the context of the MT System, a proposal has been made for “item selection” as a means of achieving a narrowing down of those variables that prove useful for recognition. A detailed discussion of this approach has been given in subsection 2.2.4. The orthogonal array is used in item selection, but it is known that, among many cases of variegated types, certain cases have proven not necessarily successful. This section takes a look at some aspects of such cases as well as some issues involving the orthogonal array and discusses how MT System and artificial intelligence relate to each other in view of the high hopes being held for the future of such interaction.

E.1 EXAMPLE OF A CUTTING VIBRATION WAVEFORM AND ITEM SELECTION

Using the example of abnormality monitoring on a cutting vibration waveform (*cf*. Section 6.1), we show how in some cases item selection performed with the help of an orthogonal array does not yield satisfactory results.

For purposes of discussing the cutting vibration waveform, we defined 20 reference lines to extract variation values and abundance values from the waveform. For these purposes, we considered the waveform symmetrical near the x axis and, as shown in Figure E.1 (a), defined reference lines with respect to the upper half of the waveform. As stated in Section 6.1, the definition of these 20 reference lines confirmed the evaluation results to be consistent with the judgment of skilled operators, but there was room for further enhancing the judgment accuracy (SN ratio) by redefining the position of the reference lines using an improved scheme.

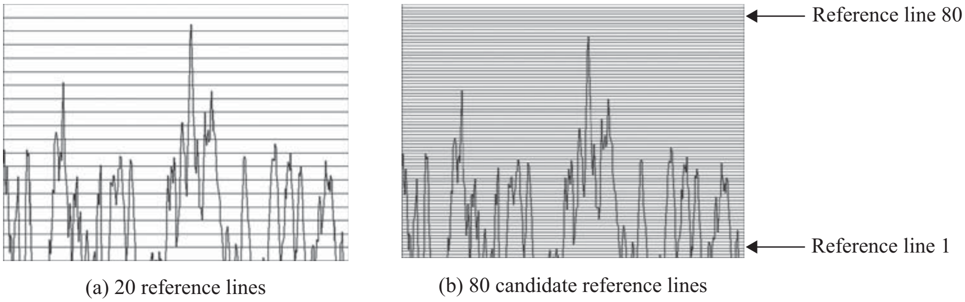


Figure E.1. Defining reference lines in search of optimum conditions.

We then proceeded to perform item selection with the use of the orthogonal array. We defined 80 candidate reference lines, as shown in Figure E.1 (b). The idea was to position these 80 reference lines as items and subsequently choose 20 that would be effective in the detection of abnormalities. There was reason, then, to expect that an improved SN ratio would be obtained.

With the reference lines numbered 1–80 ascending from the bottom line to the top line, we performed item selection with the help of an orthogonal array of the L_{128} type. The result of the analysis is displayed in Figure E.2.

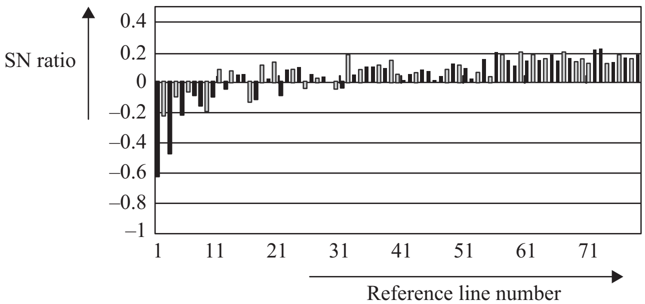


Figure E.2. Results of item selection.

Figure E.2 shows that the reference lines in the 1–10 range have effects that come out negative, and that it is past line No. 30 that the effects of the reference lines begin to show themselves to be substantial. The 20 uppermost lines, which assumed large SN ratio values, were selected so that how reference lines behave could be clearly seen, and the pattern that resulted is shown in Figure E.3.

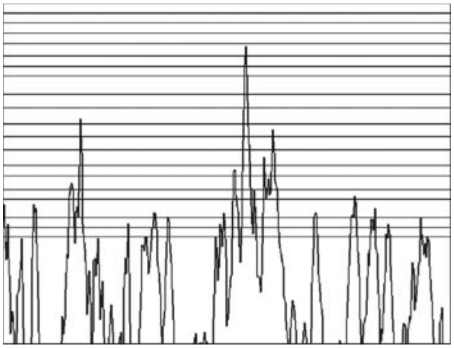


Figure E.3. Reference line redefined based on results of item selection.

The figure helps us understand that no reference lines are defined where only small vibration amplitudes occur. A computation performed on such stretches for the SN ratio of abnormal waveform detection yielded 3.18 (db). When the 20 lines were defined at an equal interval, the reading was 3.90 (db), which means that the SN ratio fell, according to Figure E.3. In case after case, item selection with the use of an orthogonal array was confirmed to be effective, but, as we have just seen, exceptions do occur.

E.2 APPLICATION OF THE GENETIC ALGORITHM

As a technique for identifying an appropriate solution from among a very large number of combinations, the *genetic algorithm* (GA) is well-known. So we apply it here for an optimal placement of reference lines. GA is an optimization theory based on the principle of the survival of the fittest, and provides a computer platform in which artificially defined genes live and thrive by repeated crossing-over, natural selection, mutation, and other processes to acquire indefinitely better-performing abilities. In industrial and other areas, GA has a track record as a tool for finding a *quasi-optimum*, if not an alternative optimum, solution.

Because scores are notated in the MT System in SN ratio terms, GA can help identify the reference line placement with the highest SN ratio from among a large number of combinations. The placement found using GA this way is shown in Figure E.4. As the figure indicates, the density of the reference lines varies depending on the position in the vertical axis. Also, unlike results obtained from the use of the orthogonal array, here, reference lines are defined even in places where the vibration is of small amplitude. Also, the SN ratio in this case is 5.3 (db), 1.4 (db) better than in the case in which 20 reference lines were defined at a regular interval.

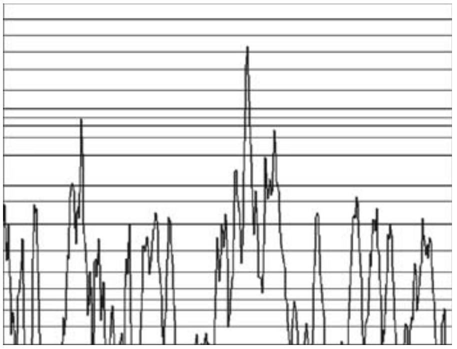


Figure E.4. Reference line placement obtained by genetic algorithm.

The genetic algorithm is one of a number of artificial intelligence theories. As an application integrated with MT System, it can achieve optimization in a more optimized manner. Another solution, the Monte Carlo Method, is said to be promising and deserves to be considered as an effective tool.

APPENDIX F

MAHALANOBIS DISTANCE COMPUTATION USING MICROSOFT EXCEL

It is virtually impossible to perform Mahalanobis Distance computation without the use of a personal computer. Dedicated analysis software products are commercially available, but Microsoft Excel can perform Mahalanobis Distance computation with comparable competence. This Section demonstrates how, using the data in Table 4.1.3 in Section 4.1, the Mahalanobis Distance computation is performed with the help of Excel.

The normalized data in Table 4.1.3 is reorganized into a table that is 14 columns by 16 rows, given that the number of items is 14 and the number of samples, 16. This is displayed in Table F.1.

[Step 1] Creation of a Correlation Matrix

A correlation matrix is easy to create with the use of [Data Analysis]. From the main Tool Bar, follow the sequence [Tools] > [Data Analysis] > [Correlation], and pull up the Correlation screen (Figure F.1). If [Data Analysis] is not found in the list of [Tools], follow the [Tools] > [Add-Ins] thread to make sure that a check mark is entered for the [Analysis ToolPak] box. If not, check the box and follow the instructions for adding the feature from the Excel installation menu.

Input range	> $B$1:$O$17
Data direction	> raw
Output range	> $A$19 (for example)

Figure F.1. Specifying the computation range.

Table F.1. Matrix of normalized data (partial reproduction of Table 4.1.3)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	1	−0.48	−0.72	−0.11	0.00	−0.82	−1.07	−2.65	0.62	−0.77	−0.59	1.61	−0.38	−0.88	2.04
3	2	2.08	−0.72	−0.11	0.00	−0.82	−1.07	0.38	−0.80	−0.77	−0.50	1.61	−0.38	−0.88	0.62
4	3	−0.48	−0.72	−0.11	0.00	−0.82	1.07	0.38	0.62	−0.77	−0.5S	1.61	−0.38	1.13	−0.80
5	4	−0.48	0.92	−0.11	0.00	−0.82	−1.07	0.38	0.62	1.29	−0.59	0.69	−0.38	−0.88	0.62
6	5	−0.48	−0.72	−0.11	0.00	−0.82	1.07	0.38	0.62	−0.77	−0.59	0.69	−0.38	1.13	−0.80
7	6	−0.48	0.92	−0.11	0.00	0.82	1.07	0.38	0.62	1.29	−0.59	−0.23	−0.38	1.13	−0.80
8	7	−0.48	−0.72	−0.11	−1.63	2.45	1.07	0.38	−0.80	−0.77	1.31	−1.15	2.65	1.13	−0.80
9	8	−0.48	−0.72	1.91	−1.63	−0.82	1.07	0.38	0.62	−0.77	3.20	−1.15	−0.38	1.13	−0.80
10	9	−0.48	−0.72	−0.11	0.00	0.82	1.07	0.38	0.62	−0.77	−0.59	−1.15	−0.38	1.13	−0.80
11	10	−0.48	0.92	−0.11	0.00	0.82	1.07	0.38	0.62	1.29	0.36	−1.15	−0.38	1.13	−0.80
12	11	−0.48	0.92	−0.11	0.00	0.82	−1.07	0.38	−0.80	1.2Э	−0.50	−0.23	−0.38	−0.88	0.62
13	12	−0.48	−0.72	3.49	3.27	0.82	0.00	0.38	0.62	−0.77	0.36	−0.23	−0.38	−0.88	0.62
14	13	−0.48	−0.72	−0.11	0.00	−0.82	−1.07	0.38	0.62	−0.77	0.36	−0.23	−0.38	−0.88	−0.80
15	14	−0.48	0.92	−0.11	0.00	0.82	0.00	0.38	0.62	1.29	0.36	−1.15	−0.38	−0.88	−0.80
16	15	2.08	−0.72	−0.11	0.00	−0.82	−1.07	−2.65	−2.22	−0.77	−0.59	−0.23	2.65	−0.88	2.04
17	16	2.08	2.56	−0.11	0.00	−0.82	−1.07	0.38	−2.22	1.29	−0.59	0.69	−0.38	−0.88	0.62

On the correlation page, click "Input Range" and drag B1–O17 to enclose this into the highlighted area; check the item [Labels in First Row]; then choose A19, for example, for [Output Range]; and finally, click [OK].

At this point, a correlation matrix in which the items number 14, with 1 added to the diagonal line of the lower triangle element, as shown in Table F.2, will be outputted.

Next, drag this matrix (A19–O33) to enclose it; follow the thread of choices: [Edit] > [Copy] > [Paste Special]; check [Transpose] in the dialog box; and copy B36 to the top. You will obtain an upper triangle element (Table F.3), which will be a transposed version of Table F.2.

From these two matrices, a correlation matrix similar to the one shown in Table 4.1.2 is created. This is accomplished, for example, by putting, Tables F.2 and F.3 together and placing the composite table over the area covered by B52–O65. At this point, the diagonal element will become 2; correct it to read 1. The correlation matrix will then look like Table F.4.

In creating a correlation matrix using the Data Analysis Tool, note that, if you are making changes to the existing normalized data, no automatic re-computation will be performed on the correlation matrix. A new correlation matrix must be created each time the existing normalized data is changed.

[Step 2] Computation of an Inverse Matrix

An inverse matrix is created using the following procedure:

(1) Specify the range to be included in the creation of the inverse matrix. Start by dragging, say, B68–O81 and enclosing it to acquire it as the range for finding the inverse matrix.
(2) Input the variables for finding the inverse matrix and specify the range for the original matrix. Drag and specify area B52–O75, as a range carved out of the existing correlation matrix, and, with it reversed and grayed out, call up the mathematical function MINVERSE.
(3) Range copy function:

Clicking the OK button while holding down the CTRL and SHIFT keys will display the inverse matrix shown in Table F.5.

[Step 3] Computation of Mahalanobis Distance

The multiplication of matrices for finding the Mahalanobis Distance is performed according to the following procedure:

(1) Row vectorization of normalized data:

In Table F.1, the data of sample No. 1 is B2–O2. To make Excel recognize this data as a row vector, drag and enclose B83–O83, and after inputting = (B2:O2) hit the ENTER key while holding down the CTRL and SHIFT keys.

(2) Computation of the Product of the Row Vector and Inverse Matrix:

Computation is performed at this point to get the product of the created row vector and the inverse matrix of the correlation matrix. To secure the range for product computation, (14 cells in the horizontal direction, such as B83–O83), drag and box off this area. Next, input =MMULT(B83:O83,$BS68:$SO$81) and simultaneously hold down the ENTER, CTRL, and SHIFT keys. Then, while leaving B83–O83 grayed out, copy, on the basis of this, what corresponds to 15 remaining pieces of data. The results of the computation are shown in Table F.6.

Table F.2. Lower triangle element of the correlation matrix

	A	B	C	D	E	h	G	H	I	J	K	L	M	N	O
19		1	2	3	4	5	6	7	8	9	10	11	12	13	14
20	1	1.00													
21	2	0.18	1.00												
22	3	−0.05	−0.08	1.00											
23	4	0.00	0.00	0.92	1.00										
24	5	−0.39	0.08	0.28	0.00	1.00									
25	6	−0.51	−0.22	−0.12	−0.22	0.44	1.00								
26	7	−0.30	0.27	0.04	0.00	0.31	0.40	1.00							
27	8	−0.84	−0.28	0.07	0.14	0.07	0.47	0.30	1.00						
28	9	−0.04	0.93	−0.09	0.00	0.21	−0.14	0.29	−0.07	1.00					
29	10	−0.29	−0.23	−0.28	−0.39	0.19	0.38	0.22	0.20	−0.21	1.00				
30	11	0.33	−0.07	0.08	0.19	−0.66	−0.43	−0.26	−0.10	−0.18	−0.57	1.00			
31	12	0.30	−0.27	−0.04	−0.31	0.31	0.00	−0.43	−0.57	−0.29	0.13	−0.26	1.00		
32	13	−0.42	−0.22	−0.33	−0.41	0.31	0.94	0.33	0.37	−0.16	0.31	−0.32	0.05	1.00	
33	14	0.53	0.01	0.23	0.29	−0.36	−0.76	−0.77	−0.51	−0.07	−0.39	0.47	0.23	−0.70	1.00

Table F.3. Upper triangle element of the correlation matrix

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
35		1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	1100	12.00	13.00	14.00
36	1	1.00	0.18	−0.05	0.00	−0.39	−0.51	−0.30	−0.84	−0.04	−0.29	0.33	0.30	−0.42	0.53
37	2		1.00	−0.08	0.00	0.08	−0.22	0.27	−0.28	0.93	−0.23	−0.07	−0.27	−0.22	0.01
38	3			1.00	0.92	0.28	−0.12	0.04	0.07	−0.09	−0.28	0.08	−0.04	−0.33	0.23
39	4				1.00	0.00	−0.22	0.00	0.14	0.00	−0.39	0.19	−0.31	−0.41	0.29
40	5					1.00	0.44	0.31	0.07	0.21	0.19	−0.66	0.31	0.31	−0.36
41	6						1.00	0.40	0.47	−0.14	0.38	−0.43	0.00	0.94	−0.76
42	7							1.00	0.30	0.29	0.22	−0.26	−0.43	0.33	−0.77
43	8								1.00	−0.07	0.20	−0.10	−0.57	0.37	−0.51
44	9									1.00	−0.21	−0.18	−0.29	−0.16	−0.07
45	10										1.00	−0.57	0.13	0.31	−0.39
46	11											1.00	−0.26	−0.32	0.47
47	12												1.00	0.05	0.23
48	13													1.00	−0.70
49	14														1.00

Table F.4. Results of correlation matrix computation

51		1	2	3	4	5	6	7	8	9	10	11	12	13	14
52	1	1.00	0.18	−0.05	0.00	−0.39	−0.51	−0.30	−0.84	−0.04	−0. 9	0.33	0.30	−0.42	0.53
53	2	0. 18	1.00	−0.08	0.00	0.08	−0.22	0.27	−0.28	0.83	−0.23	−0.07	−0.27	−0.22	0.01
54	3	−0.05	−0.08	1.00	0.92	0.28	−0.12	0.04	0.07	−0.08	−0.28	0.08	−0.04	−0.33	0.23
55	4	0.00	0.00	0.92	1.00	0.00	−0.22	0.00	0. 14	0.00	−0.39	0. 19	−0.31	−0.41	0.29
56	5	−0.39	0.08	0.28	0.00	1.00	0.44	0.31	0.07	0.21	0.19	−0.88	0.31	0.31	−0.36
57	6	−0.51	−0.22	−0.12	−0.22	0.44	1.00	0.40	0.47	−0.14	0.38	−0.43	0.00	0.94	−0.76
58	7	−0.30	0.27	0.04	0.00	0.31	0.40	1.00	0.30	0.29	0.22	−0.28	−0.43	0.33	−0.77
59	8	−0.84	−0.28	0.07	0.14	0.07	0.47	0.30	1.00	−0.07	0.20	−0. 10	−0.57	0.37	−0.51
60	9	−0.04	0.93	−0.09	0.00	0.21	−0.14	0.29	−0.07	1.00	−0.21	−0. 19	−0.29	−0.16	−0.07
61	10	−0.29	−0.23	−0.28	−0.99	0.19	0.39	0.22	0.20	−0.21	1.00	−0.57	0. 19	0.31	−0.39
62	11	0.33	−0.07	0.08	0.19	−0.66	−0.43	−0.26	−0.10	−0.18	−0.57	1.00	−0.26	−0.32	0.47
63	12	0.30	−0.27	−0.04	−0.31	0.31	0.00	−0.43	−0.57	−0.23	0.13	−0.26	1.00	0.05	0.23
64	13	−0.42	−0.22	−0.33	−0.41	0.31	0.34	0.33	0.37	−0.16	0.31	−0.32	0.05	1.00	−0.70
65	14	0.53	0.01	0.23	0.23	−0.36	−0.76	−0.77	−0.51	−0.07	−0.33	0.47	0.23	−0.70	1.00

Table F.5. Inverse matrix of the correlation matrix, as created

67		1	2	3	4	5	6	7	8	9	10	11	12	13	14
68	1	7.01	−1.66	8.24	−7.89	−0.24	−2.92	−1.42	5.50	2.65	0.00	−1.70	−1.81	3.10	−0.27
69	2	−1.66	134.89	−246.47	235.95	82.75	−41.79	42.48	75.80	−131.04	17.11	49.12	102.41	14.51	18.52
70	3	8.24	−246.47	534.83	−516.96	−178.29	88.22	−84.90	−137.55	243.06	−42.65	−112.19	−210.13	−30.86	−28.39
71	4	−7.89	235.95	−516.96	505.31	173.31	−95.47	78.01	130.83	−231.14	44.72	110.77	203.27	38.88	21.06
72	5	−0.24	82.75	−178.29	173.31	63.98	−35.39	25.44	47.25	−81.46	16.05	39.59	68.61	14.58	6.49
73	6	−2.92	−41.79	88.22	−95.47	−35.39	52.89	−6.74	−26.94	38.24	−14.01	−21.75	−36.02	−37.12	7.32
74	7	−1.42	42.48	−84.90	78.01	25.44	−6.74	22.60	26.45	−43.03	3.10	14.18	36.40	0.00	15.39
75	8	5.50	75.80	−137.55	130.83	47.25	−26.94	26.45	52.48	−72.59	9.23	26.61	60.77	11.18	13.78
76	9	2.65	−131.04	243.06	−231.14	−81.46	38.24	−43.03	−72.59	129.95	−15.41	−47.13	−99.02	−11.53	−19.74
77	10	0.00	17.11	−42.65	44.72	16.05	−14.01	3.10	9.23	−15.41	7.89	12.24	16.10	8.36	−2.64
78	11	−1.70	49.12	−112.19	110.77	39.59	−21.75	14.18	26.61	−47.13	12.24	28.50	43.73	8.65	0.51
79	12	−1.81	102.41	−210.13	203.27	68.61	−36.02	36.40	60.77	−99.02	16.10	43.73	88.64	13.13	13.48
80	13	3.10	14.51	−30.86	38.88	14.58	−37.12	0.00	11.18	−11.53	8.36	8.65	13.13	31.50	−5.59
81	14	−0.27	18.52	−28.39	21.06	6.49	7.32	15.39	13.78	−19.74	−2.64	0.51	13.48	−5.59	18.49

Table F.6. Results of computation of product of row vector and inverse matrix

83	d1	−0.48	−0.72	−0.11	0.00	−0.82	−1.07	−2.65	0.62	−0.77	−0.59	1.61	−0.38	−0.88	2.04
84	d2	2.08	−0.72	−0.11	0.00	−0.82	−1.07	0.38	−0.80	−0.77	−0.59	1.61	−0.38	−0.88	0.62
85	d3	−0.48	−0.72	−0.11	0.00	−0.82	1.07	0.38	0.62	−0.77	−0.59	1.61	−0.38	1.13	−0.80
86	d4	−0.48	0.92	−0.11	0.00	−0.82	−1.07	0.38	0.62	1.29	−0.59	0.69	−0.38	−0.88	0.62
87	d5	−0.48	−0.72	−0.11	0.00	−0.82	1.07	0.38	0.62	−0.77	−0.59	0.69	−0.33	1.13	−0.80
88	d6	−0.48	0.92	−0.11	0.00	0.82	1.07	0.38	0.62	1.29	−0.59	−0.23	−0.33	1.13	−0.80
83	d7	−0.48	−0.72	−0.11	−1.63	2.45	1.07	0.38	−0.80	−0.77	1.31	−1.15	2.65	1.13	−0.80
90	d8	−0.48	−072	−1.91	−1.63	−0.82	1.07	0.38	0.62	−0.77	3.20	−1.15	−0.33	1.13	−0.80
91	d9	−0.48	−072	−0.11	0.00	0.82	1.07	0.38	0.62	−0.77	−0.59	−1.15	−0.38	1.13	−0.80
92	d10	−0.48	0.92	−0.11	0.00	0.82	1.07	0.38	0.62	1.29	0.36	−1.15	−0.38	1.13	−0.80
93	d11	−0.48	0.92	−0.11	0.00	0.82	−1.07	0.38	−0.80	1.29	−0.59	−0.23	−0.38	−0.88	0.62
94	d12	−0.48	−0.72	3.49	3.27	0.82	0.00	0.38	0.62	−0.77	0.36	−0.23	−0.38	−0.88	0.62
95	d13	−0.48	−0.72	−0.11	0.00	−0.82	−1.07	0.38	0.62	−0.77	0.36	−0.23	−0.38	−0.88	−0.80
96	d14	−0.48	0.92	−0.11	0.00	0.82	0.00	0.38	0.62	1.29	0.36	−1.15	−0.38	−0.88	−0.80
97	d15	2.08	−0.72	−0.11	0.00	−0.82	−1.07	−2.65	−2.22	−0.77	−0.59	−0.23	2.65	−0.88	2.04
98	D16	2.08	2.56	−0.11	0.00	−0.82	−1.07	0.38	−2.22	1.29	−0.59	0.69	−0.38	−0.88	0.62

Table F.7. Results of computations of Mahalanobis Distances of unit data

117	MD1	1.0714
118	MD2	1.0714
119	MD3	0.877
120	MD4	1.0675
121	MD5	0.9722
122	MD6	0.5913
123	MD7	1.0714
124	MD8	1.0556
125	MD9	1.0556
126	MD10	0.75
127	MD11	1.0714
128	MD12	1.0675
129	MD13	1.0675
130	MD14	1.0675
131	MD15	1.0714
132	MD16	1.0714

(3) Computation of Mahalanobis Distance Using Sum of Products of Normalized Data

We will now compute the sum of products of the row vectors in Table F.6 and the normalized data in Table F.1. For the sum of products, use SUMPRODUCT for mathematical functions. First, for the data for d1, call up SUMPRODUCT, and assign B83–O83 to array 1 and normalized data for d1 B2–O2 to array 2; then click OK. And, by copying downward exactly from product d2 to product d16, you can completely compute the Mahalanobis Distance of the Unit Data (Table F.7).

If you wish to try out an analysis of the data, etc. shown in the examples given above on your own PC, you can perform Mahalanobis Distance computations using the method discussed. However, if there is a change in the number of items of Signal Data, or if you wish to perform item selection, the same procedure discussed above will have to be repeated many times, and it will turn out to be a more tedious operation than one imagines. For purposes of performing various analyses, it will be more convenient to utilize commercially available dedicated software products.

APPENDIX G

Paley's Construct for Generation of Hadamard Matrice[10,12]

MT System uses the orthogonal array of the two-level type when working on item selection, cause diagnostics, etc. To serve general test designing purposes, orthogonal arrays of the "power-raised" type, the sizes of which are represented in terms of 2^n, as with L_8, L_{16}, etc., are used in many cases. With orthogonal arrays of this type, the existence of interaction between items will make the effect evident in specific columns in the form of effect confounding, making it impossible to tell if the observed effect is that of the item allotted to the column or that of interaction with another item allotted to another column.

Thus, when item selection is undertaken under MT System, if interaction between, or among, items is present, an orthogonal array is commonly used that effectively scatters the effect of the interaction evenly among all columns. Paley's cyclic orthogonal array represents an example of such a type of array.

Paley's cyclic orthogonal array, proposed by R.E.A.C. Paley in 1933, is a matrix created using what is known as "quadratic residue." Because it is relatively easy to generate it and the order of power does not necessarily come in the form of 2^n, it lends itself to a wide variety of applications, in addition to item selection in the context of the MT System.

G.1 QUADRATIC RESIDUE

For the sake of discussion here, we will call p the prime of an odd number. The residue left when an integer is divided by this p will be p in number, starting with 0 (0 counting as a special residue), going on 1, 2, …, and ending with $p - 1$. Now, when the residue left after an integer is divided by p and the residue is q, the q will be expressed as q (mod p).

And if we pick out any randomly chosen one of the sequences *GF* (p) of the residue and call it a, and, again, if there is an integer y which can be expressed as:

$$y^2 = a \pmod{p},$$

then, a is said to be a quadratic residue of mod p. A few concrete examples with $p = 11$ are shown below, where, for y = 1 … 10, the following instances of the quadratic residue work out:

$$
\begin{aligned}
\text{When } y=1&: y^2=1 \rightarrow 1 \pmod{11}\\
y=2&: y^2=4 \rightarrow 4 \pmod{11}\\
y=3&: y^2=9 \rightarrow 9 \pmod{11}\\
y=4&: y^2=16 \rightarrow 5 \pmod{11}\\
y=5&: y^2=25 \rightarrow 3 \pmod{11}\\
y=6&: y^2=36 \rightarrow 3 \pmod{11}\\
y=7&: y^2=49 \rightarrow 5 \pmod{11}\\
y=8&: y^2=64 \rightarrow 9 \pmod{11}\\
y=9&: y^2=81 \rightarrow 4 \pmod{11}\\
y=10&: y^2=100 \rightarrow 1 \pmod{11}
\end{aligned}
$$

In other words, when y is 1 … 10, the quadratic residues of *GF* (1) are five – 1, 3, 4, 5 and 9. For when $y = 1 \ldots \infty$ as well, the quadratic residues for *GF* (11) will be the same set of five numbers 1, 3, 4, 5 and 9. Generally speaking, corresponding to p number of elements in *GF* (p), there exist $(p - 1)/2$ number of quadratic residues, where the remaining $(p - 1)/2$ number of elements are *quadratic non-residues*, of which there are as many as there are quadratic residues. This, then, is made use of, as shown in the next section, toward the "+1 or –1" determination of the elements of the Hadamard matrix.

G.2 GENERATION OF PALEY'S CYCLIC MATRIX

We will now consider the following row vectors as they concern p number of numbers $a = 0, 1, 2, \ldots, p - 1$, with χ (a), called the Legendre function defined as follows:

$$
\begin{aligned}
\chi(a) &= 0 \ (a=0)\\
&= +1 \ (\text{quadratic residue mod } (p))\\
&= -1 \ (\text{quadratic non-residue mod } (p)).
\end{aligned}
$$

Let's look at a concrete example, again using the $p = 11$ that we used a little earlier. Given that the quadratic residues of GF (11) are 1, 3, 4, 5 and 9, we will get:

a	0	1	2	3	4	5	6	7	8	9	10
$\chi(a)$	0	+1	−1	+1	+1	+1	−1	−1	−1	+1	−1

We will set up $\chi(a)$ as the first row, as shown below, and, for the second and following rows, shift the entries to the right by one grid, repeating the same procedure for each of the remaining rows, such that, after the procedure has been repeated p times, a cyclic matrix Q can be assumed (Table G.1).

Table G.1. Cyclic matrix Q as of 11th round

$$Q = \begin{pmatrix} 0 & +1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 & -1 \\ -1 & 0 & +1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 \\ +1 & -1 & 0 & +1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 \\ -1 & +1 & -1 & 0 & +1 & -1 & +1 & +1 & +1 & -1 & -1 \\ -1 & -1 & 1 & -1 & 0 & +1 & -1 & +1 & +1 & +1 & -1 \\ -1 & -1 & -1 & +1 & -1 & 0 & +1 & -1 & +1 & +1 & +1 \\ +1 & -1 & -1 & -1 & +1 & -1 & 0 & +1 & -1 & +1 & +1 \\ +1 & +1 & -1 & -1 & -1 & +1 & -1 & 0 & +1 & -1 & +1 \\ +1 & +1 & +1 & -1 & -1 & -1 & +1 & -1 & 0 & +1 & -1 \\ -1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 & -1 & 0 & +1 \\ +1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 & -1 & 0 \end{pmatrix}$$

If we create at this point the following matrix as of the (p + 1)th round, matrix H_n is an orthogonalized Hadamard matrix:

$$H_n = \begin{pmatrix} 1 & -1\frac{T}{p} \\ 1_p & Q + I_p \end{pmatrix}$$

Here 1_p refers to a column vector with p number of ls lined up, while -1_p^T refers to a transposed column vector (i.e., row vector) with p number of –ls lined up. Moreover, I_p shows a unit matrix (a matrix in which only the diagonal is 1 [one] and the other components are zeroes) as of the p-th round.

From this point on, using the cyclic matrix Q created above, the Hadamard matrix as of the 12th round can be created as follows:

Table G.2. Paley's cyclic Hadamard matrix H_{12}

$$H_{12} = \begin{pmatrix} +1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ +1 & +1 & +1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 & -1 \\ +1 & 1 & +1 & +1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 \\ +1 & +1 & -1 & +1 & +1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 \\ +1 & -1 & +1 & -1 & +1 & +1 & -1 & +1 & +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 & +1 & +1 & +1 & -1 \\ +1 & -1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 & +1 & +1 & +1 \\ +1 & +1 & -1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 & +1 & +1 \\ +1 & +1 & +1 & -1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 & +1 \\ +1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 \\ +1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 & -1 & +1 & +1 \\ +1 & +1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 & -1 & +1 \end{pmatrix}$$

The "0s" in Table G.1 are replaced by "+1s" in Table G.2. The reader, on his or her account, should confirm with respect to this Paley's cyclic Hadamard matrix H_{12} just examined that each column in it contains six positive and negative signs and that any arbitrarily chosen pair of columns are orthogonally bound. (Note that the orthogonal array L_{12}, which can be created from the H_{12} generated here, is different from the orthogonal array L_{12} in common use in connection with Quality Engineering.)

The cyclic matrix Q' encircled by a broken line in Table G.2 is called "Paley's cyclic matrix," and with its use an alternative solution to H_{p+1} can be generated as shown Table G3.

Table G.3. Alternative solution to Paley's cyclic Hadamard matrix

Solution (1)

$$\begin{pmatrix} +1 & +1 & +1 & +1 \\ +1 & & & \\ +1 & & -Q' & \\ +1 & & & \end{pmatrix}$$

Solution (2)

$$\begin{pmatrix} +1 & +1 & +1 & +1 \\ +1 & & & \\ +1 & & -Q'^{T} & \\ +1 & & & \end{pmatrix}$$

Solution (3)

$$\begin{pmatrix} +1 & -1 & -1 & -1 \\ +1 & & & \\ +1 & & Q'^{T} & \\ +1 & & & \end{pmatrix}$$

As seen here, when p is set up as an odd prime using Paley's method, a cyclic Hadamard matrix as of the $(p+1)$th round can be generated.

Note in passing that primes, 1, 2, 3, 5, 7, 11, 13, 17, 19, etc., when divided by 4, in most cases produce a residue of 3 (or 1 in some cases). Thus, when adding 1 to a prime, one obtains a multiple of 4, and that is why the Hadamard matrix of this type is also referred to as a matrix of the $4 \times n$ type.

If we delete the first column of the cyclic Hadamard matrix that we just created, we have the so-called orthogonal array. With an orthogonal array of the $4 \times n$ type, the interaction between any arbitrarily chosen pair of columns is dispersed evenly to the other columns. This is easily inferred given that the matrix is created by shifting all rows by one column in a staggered fashion. In view of these considerations, use of the orthogonal array of the $4 \times n$ type is recommended in the interest of avoiding the interaction issue when the MT System is used for item selection.

BIBLIOGRAPHY AND REFERENCE SOURCES

This Section lists related books, explanatory textbooks, and other reference literature that has been published on the subject of the MT System.

The reference literature is for the most part Japanese-language publications. There are other publications that should also be included in this category, but for present purposes inclusion had to be limited to representative works.

BIBLIOGRAPHY (IN ENGLISH)

Genichi Taguchi et al. (2005): Taguchi's Quality Engineering Handbook. John Wiley & Sons, Inc.
Outline: A handbook published in the United States with a substantial number of pages dedicated to the MT System. Introduces comments by Dr. Taguchi and a number of cases.

Genichi Taguchi et al. (2001): The Mahalanobis-Taguchi System. McGraw-Hill.
Outline: An introduction to the MT System (MT Method) explaining it with references to concrete examples, taken mostly from case studies in Japan.

Genichi Taguchi and Rajesh Jugulum (2002): The Mahalanobis-Taguchi Strategy. John Wiley & Sons, Inc.
Outline: A summing-up by Dr. Jugulum, who pursues studies in quality engineering in the United States, of his work with and under the guidance of Dr. Taguchi. Explains the MT System and Gram-Schmidt's computation methodology citing cases based on medical and other data.

BIBLIOGRAPHY (IN JAPANESE)

Genichi Taguchi et al. (2002): Course in Quality Engineering Application—Technical Development within the Framework of the MT System, Japanese Standards Association.
Outline: Includes commentary by Dr. Taguchi and introduces the work of many researchers. Author's commentary includes discussion of the MT Method, MTA Method, and Schmidt's computation methodology, which assigns signs to distances. Thirty-one cases introduced provide substantial food for thought.

Yoshiko Hasegawa (2004): Introduction to the Mahalanobis-Taguchi (MT) System, JUSE Press.
Outline: A primer designed to present in a reader-friendly manner an overview of the MT System and the fields in which its applications are in common use. Consists of a Basics and Essentials section, an Applica-

tions section, and a Q&A section. Illustrative examples are drawn from real-life situations such as medical issues, all accompanied by thoughtful running commentaries.

Kazuo Tatebayashi (2004): A Taguchi Method Primer. JUSE Press.

Outline: Discusses quality engineering addressing among its readers, beginners and technicians who have cultivated familiarity with statistical methodology as well. In terms of the MT System, the book explains basic subjects including computation methods.

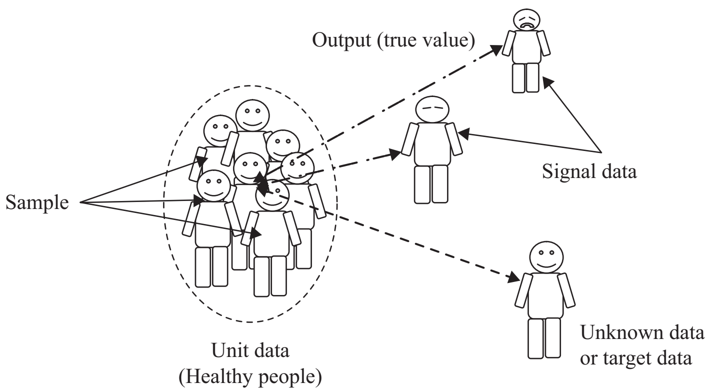
REFERENCES

1. Kanetaka, Tatsuji. 1987. Experiment design for medical treatment. Diagnosis of special examination cases with Mahalanobis's generalized distance. *Standardization and Quality Control* 40 (10): 57–64.
2. Taguchi, Genichi. 1995. Quality engineering for pattern recognition. *Journal of Quality Engineering Forum* 3 (2): 2–5.
3. Taguchi, Genichi. 2005. Objective function and generic function (6)—Prediction by Taguchi methods. *Journal of Quality Engineering Forum* 13 (3): 5–10.
4. Fujita, Ichiro, Keiji Tanaka, Minami Ito, and Kang Cheng. 1992. Columns for visual features of objects in monkey inferotemporal cortex. *Nature* 360: 343–46. DOI: 10.1038/360343a0
5. Taguchi, Genichi. 1997. Method for pattern recognition. US Patent 5,684,892, filed Aug. 22, 1995, and issued Nov. 4.
6. Hasegawa, Yoshiko. 1997. How to treat missing data in health examination using Mahalanobis distance. Proceedings for *3rd Annual International Total Product Development Symposium*. ed. Subir Chowdhury, 619–630, Livonia, MI: American Supplier Institute.
7. Teshima, Shoichi, Okuno Taku, Aono Fumiaki, Sato Koichi, and Kubo Hiroshi. 1999. A study on cutting state observation using Mahalanobis distance. *Journal of the Japan Society for Precision Engineering* 65 (9): 1325–29.
8. Teshima, Shoichi, Bando Tomonori, and Jin Dan 1998. Research on defect detection using the Mahalanobis-Taguchi System Method. In *4th Annual International TPD Symposium: Robust Engineering Conference*, ed. Subir Chowdhury, 170–80. Livonia, MI: American Supplier Institute.
9. Guyatt, Gordon, et al. 1992. Evidence-based medicine: A new approach to teaching the practice of medicine. *JAMA* 268: 2420–5. DOI: 10.1001/jama.1992.03490170092032
10. Hedayat, A. S., Neil J. A. Sloane, and John Stufken. 1999. *Orthogonal arrays: Theory and applications*. New York: Springer.
11. Yoshino, Sohei, Yano Koya, Ishii Chiharu, and Wada Tadashi. 2006. Estimation of real estate value by MT System (3)—Price forecasting from single-data unit space data without Schmidt orthogonal estimation. *Journal of Quality Engineering Forum* 14 (1): 68–76.
12. Kiyasu, Zenichi. (1980): Hadamard procession and it's applications. The Institute of Electronics and Communication Engineers.

Glossary

DEFINITION OF TERMS

It is an established practice for books to define the terminology used in the opening pages. Terms used in MT System discussions reflect those defined by the Quality Engineering Society,[16,17] but they are not quite as up to date as they would be expected to be. If we were to have defined the relevant terms at the beginning of the book, we were afraid that they might very well prove rather confusing, so we decided to defer the glossary to this chapter on the assumption that by now the reader will have acquired a good grasp of the overall picture.



General concept of terms

Unit Space — A population that constitutes the "center/reference" for pattern recognition and is expected to be homogeneous with respect to the objective. If the context were one accommodating physical examinations, the Unit Space would consist of "healthy people." It is sometimes also referred to as "reference space" or "normal space," names that in certain respects help one to understand the concept better, but, in light of the fact that we are dealing here with the MT System as a measuring technique, "Unit Space" is used as the defining term.

Sample of a Unit Space	A formative element of a Unit Space. If the Unit Space is to be made up of healthy people, it refers to individual people in good health like, say, Mr. A, Ms. B, etc. *Though Taguchi uses "Member" in this case, this book uses "Sample" because the latter is more familiar.*
Unit Space data	Unit Space attribute value or variable (item) value. For example, body temperatures, blood sugar values, etc. collected from healthy people.
Signal Data	Samples not included in a Unit Space, or their attributes. Used in T Method-1. At the time the MT Method was originally proposed, unknown data was included in what was generally referred to as "Signal Data," but after the birth of the T Method the term began to be used mainly to refer to data outside the Unit Space.
Output value	True value, in relation to the object of the Unit Data and Signal Data used in T Method-1. Corresponds to the objective variable in multivariate analysis. Illness recovery time, real estate prices, etc., the various physical attribute values that we wish to predict are all examples of output values.
Unknown (target) data	Data for which the Mahalanobis distance or output value is unknown.

INDEX