

CHAPTER 5

T METHOD APPLICATION PROCEDURES AND KEY POINTS

In this Chapter, application procedures for T Method (1) and RT Method (=T Method (3)) are explained.

T Method (1) is discussed using “yield prediction in the context of manufacturing production processes” as an illustrative example. It concerns the prediction of a production yield on the basis of process data acquired from day-to-day manufacturing routines such as temperature, pressure, and other readings.

RT Method is discussed using a numeral pattern as an illustrative example. Unlike the case given in Chapter 4, density information (1,0) of a pixel for use *in raw form* is utilized. Consequently, a Unit Space is created for the numeral “7” as part of the procedure for determining whether or not the object pattern can be regarded as belonging in the “5” group.

5.1 YIELD PREDICTION FOR MANUFACTURING-PRODUCTION USING T METHOD-1

In most cases in an ongoing manufacturing process, it is difficult to interrupt and insert tests into intervals of production runs. If daily manufacturing data or data pertaining to manufacturing under altered conditions is available, use of T Method-1 may make it possible to identify the conditions for improvement.

Process data for the case given here is shown in Table 5.1. In this table, the yields, which are the output, are rearranged in ascending order of the value of output. There are six items: B temperature (°C), C temperature (°C), Pressure 1 (10×MPa), Pressure 2 (10×MPa), preheating time (hours), and manufacturing time (minutes). These will all turn out as measured values. Since the yield, which comes out as an output value, is data that can assume any value between 0.0 (%) and 100.0 (%), that is to say, any percentage value, the data should be analyzed upon its omega conversion

(logit transformation). However, given that the yield has proven to be in this case less than 90%, the output value will be analyzed in raw data form. (For a detailed discussion on omega transformation, see Note 5.1.)

Table 5.1. Process data

Data No.	“B” temp	“C” temp	P 1	P 2	Pre-ht time	Manuf time	Yield
1	575.0	274.0	205.5	210.5	5.0	180.0	81.55%
2	575.0	304.0	207.5	218.5	6.0	120.0	82.99%
3	570.0	279.0	199.5	207.5	7.0	120.0	83.03%
4	577.0	205.0	173.0	164.0	7.0	120.0	84.56%
5	573.0	254.0	160.0	164.0	7.0	120.0	84.60%
6	577.0	216.0	176.0	168.0	6.0	120.0	85.52%
7	582.0	222.5	171.0	167.0	5.0	180.0	89.47%

5.1.1 DEFINITION OF THE UNIT SPACE

Figure 5.1 shows the distribution of outputs, that is, yields, displayed in Table 5.1. The figure allows us to judge that a concentration of high yields occurs within the 84.0–85.0% yield range. The total number of data entries is small (seven), and even at the spot of highest data concentration the yield is not 100%. For this reason, it is appropriate for the purposes of yield prediction to use T Method-1.

Given the yield distribution, 4 and 5, two samples found in the 84.0–85.0% yield range, are selected from Table 5.1 as samples of the Unit Space. This data is, again, shown in Table 5.2.

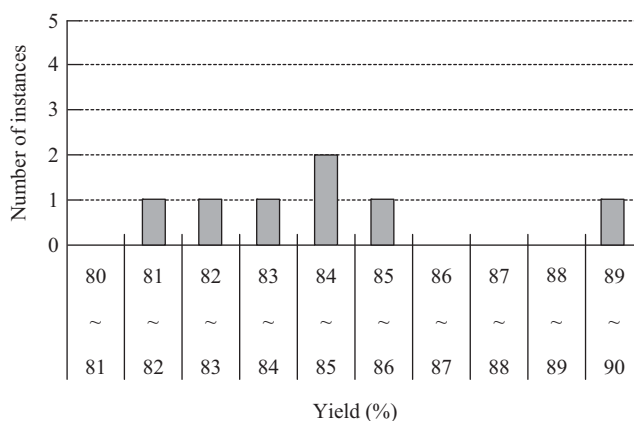


Figure 5.1. Distribution of outputs.

Table 5.2. Data in the unit space and the average values

Sample No.	“B” temp	“C” temp	P 1	P 2	Pre-ht time	Manuf time	Yield
4	577.0	205.0	173.0	164.0	7.0	120.0	84.56%
5	573.0	254.0	160.0	164.0	7.0	120.0	84.60%
Average	575.0	229.5	166.5	164.0	7.0	120.0	84.58%

We will next find from the two samples of the Unit Space the average value \bar{x}_j ($j = 1, 2, \dots, 6$) of each item and the average value M_θ of the output value (yield). The average values are as shown in the bottom row of Table 5.2, in T Method-1; it is these average values that form the center of the Unit Space.

5.1.2 DEFINITION OF SIGNAL DATA

Of the process data tabulated in Table 5.1, the five data entries, Nos. 1–3 and Nos. 6 and 7, that were not selected for membership in the Unit Space as samples, are treated as Signal Data, as shown in Table 5.3.

Table 5.3. Signal data

Data No.	“B” temp	“C” temp	P 1	P 2	Pre-ht time	Manuf time	Yield
1	575.0	274.0	205.5	210.5	5.0	180.0	81.55%
2	575.0	304.0	207.5	218.5	6.0	120.0	82.99%
3	570.0	279.0	199.5	207.5	7.0	120.0	83.03%
6	577.0	216.0	176.0	168.0	6.0	120.0	85.52%
7	582.0	222.5	171.0	167.0	5.0	180.0	89.47%

5.1.3 NORMALIZATION OF SIGNAL DATA

The average values obtained from the Unit Space samples are used for Signal Data normalization. Normalization is performed by subtracting the averages of the items and yield of the Unit Space from the yield value from each item in the Signal Data. Below is a demonstration of how “B” temperature X_{11} and output M_1 for sample No. 1 are computed:

$$X_{11} = 575.0 - 575.0 = 0.0 \quad (5.1)$$

$$M_1 = 81.55 - 84.58 = -3.03 \quad (5.2)$$

All the pieces of normalized Signal Data are tabulated in Table 5.4. In a statistics-based approach normalization is generally performed by dividing what remains after subtracting the average values from the standard deviation of the item, but T Method-1 dispenses with this computation. This is because in many cases items that have zero standard deviation in the Unit Space, which would make the computation impossible, are of key importance for prediction and estimation purposes.

Table 5.4. Normalized signal data

Data No.	“B” temp	“C” temp	P 1	P 2	Pre-ht time	Manuf time	Yield
1	0.0	44.5	39.0	46.5	−2.0	60.0	−3.03%
2	0.0	74.5	41.0	54.5	−1.0	0.0	−1.59%
3	−5.0	49.5	33.0	43.5	0.0	0.0	−1.55%
6	2.0	−13.5	9.5	4.0	−1.0	0.0	0.94%
7	7.0	−7.0	4.5	3.0	−2.0	60.0	4.89%

5.1.4 COMPUTATION OF PROPORTIONAL COEFFICIENT AND SN RATIO OF EACH ITEM OF SIGNAL DATA

To find out which of the items will be useful for prediction and estimation, an item-by-item computation of the proportional coefficient β and SN ratio η is performed. Figure 5.2 is a graph illustration of the relationships between temperature or pressure, etc., on the one hand, and the output value (yield). The horizontal axis represents the output value. For both the horizontal and the vertical axis, the zero point constitutes the center of the Unit Space.

In T Method-1, SN ratios η and proportional coefficients β are calculated from the relationship between the output value and the item value. The larger the SN ratio η here, the closer to a straight line is the relationship between the output value (yield) and the other items. In other words, such items will act as more significant contributors when a general estimation of the yield is to be made. Furthermore, proportional coefficient β will provide a measure of the steepness of the incline of the straight line for each of the items in Figure 5.2.

The following matters can be explained by the six graphs in Figure 5.2.

(a) Yield and “B” Temperature

Given the right upward incline of the line, the proportional coefficient β is positive. Also, the graph shows that the data is arranged in relatively neat formation along the regression line that passes the zero point and outbound scatter is limited. That is to say, the SN ratio η is large, leading to the assumption that “B” temperature is well-suited to the purpose of general yield estimation.

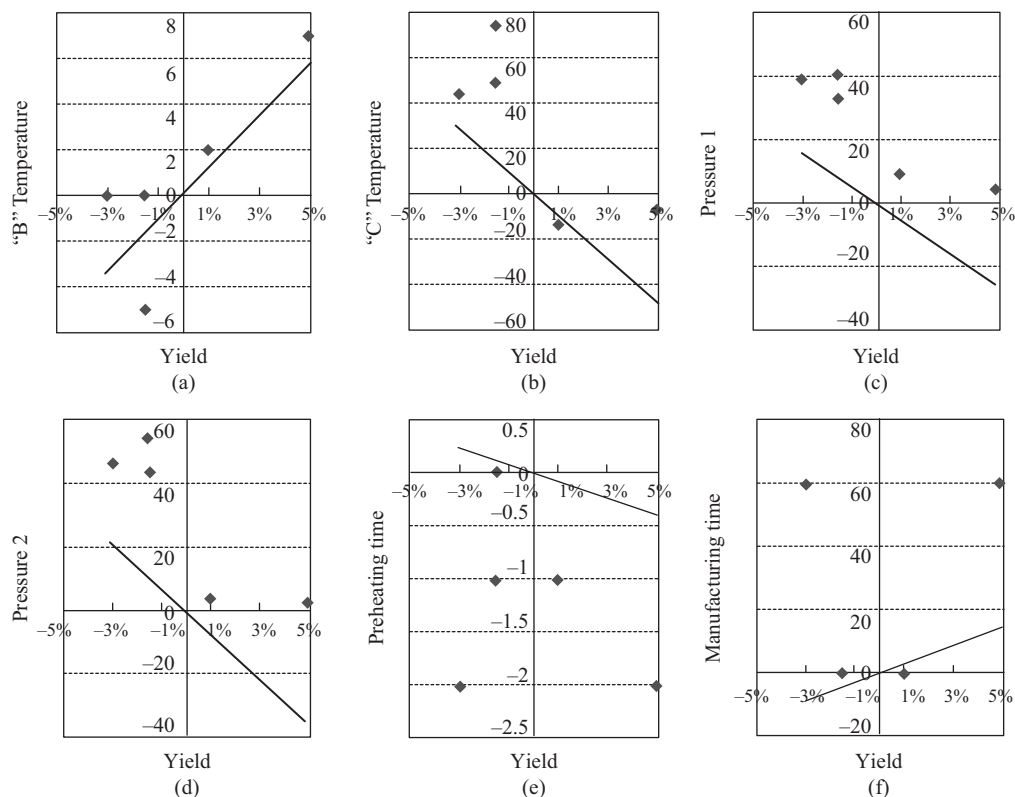


Figure 5.2. Scatter of output value yields and items (normalized data).

(b) Yield and "C" Temperature

Given the right downward incline of the line, the proportional coefficient β turns out negative. Also, the graph shows that the data is arranged in relatively neat formation along the regression line that passes the zero point and outbound scatter is limited. That is to say, the SN ratio η is large, giving credence to the assumption that "C" temperature is also well-suited to the purpose of general yield estimation.

(c) Yield and Pressure 1

The scatter of the data away from the regression line that passes the zero point is considerable. That is to say, the SN ratio η is somewhat small, so it appears that Pressure 1 is not very useful for general yield estimation.

(d) Yield and Pressure 2

The scatter of the data away from the regression line that passes the zero point is considerable. That is to say, the SN ratio η is somewhat small, so it seems that Pressure 2 is not very useful for general yield estimation.

(e) Yield and Preheating Time

The scatter of the data away from the regression line that passes the zero point is substantial. That is to say, the SN ratio η is small, so it appears that preheating time is not useful for general yield estimation.

(f) Yield and Manufacturing Time

The scatter of the data away from the regression line that passes the zero point is substantial. That is to say, the SN ratio η is small, so it seems that manufacturing time is not useful for general yield estimation.

Next, we compare the points explained in (a) to (f) above with the results of the item-by-item computation of the proportional coefficient β and the SN ratio η . The computation of the proportional coefficient β and the SN ratio η is performed item by item with the use of normalized data X_{ij} and normalized output value M_i . How to compute the first item No. 1, “B” temperature, is explained below.

Effective divider

$$\begin{aligned} r &= M_1^2 + M_2^2 + \cdots + M_l^2 \\ &= (-0.0303)^2 + (-0.0159)^2 + \cdots + 0.0489^2 = 0.00389 \end{aligned} \quad (5.3)$$

Total variation

$$\begin{aligned} S_{T1} &= X_{11}^2 + X_{21}^2 + \cdots + X_{l1}^2 \\ &= 0.0^2 + 0.0^2 + \cdots + 7.0^2 = 78.0 \quad (f=5) \end{aligned} \quad (5.4)$$

Variation of proportional term

$$\begin{aligned} S_{\beta 1} &= \frac{(M_1 X_{11} + M_2 X_{21} + \cdots + M_l X_{l1})^2}{r} \\ &= \frac{\{(-0.0303) \times 0.0 + (-0.0159) \times 0.0 + \cdots + 0.0489 \times 7.0\}^2}{0.00389} \\ &= 49.4433 \quad (f=1) \end{aligned} \quad (5.5)$$

Error variation

$$S_{e1} = S_{T1} - S_{\beta1} = 78.0 - 49.4433 = 28.5567 \quad (f=4) \quad (5.6)$$

Error variance

$$V_{e1} = \frac{S_{e1}}{l-1} = \frac{28.5567}{4} = 7.1392 \quad (5.7)$$

Therefore, the proportional coefficient β and the SN ratio η can be found as follows:
Proportional Coefficient:

$$\begin{aligned} \beta_1 &= \frac{M_1 X_{11} + M_2 X_{21} + \cdots + M_l X_{l1}}{r} \\ &= \frac{(-0.0303) \times 0.0 + (-0.0159) \times 0.0 + \cdots + 0.0489 \times 7.0}{0.00389} \\ &= 112.7296 = 112.73 \end{aligned} \quad (5.8)$$

Given the SN ratio η_1 $S_{\beta1} > V_{e1}$, the equation works out as follows. The SN ratio in this case is a duplicate ratio.

$$\eta_1 = r \frac{\frac{1}{r} (S_{\beta1} - V_{e1})}{V_{e1}} = \frac{\frac{1}{0.00389} (49.4433 - 7.1392)}{7.1392} = 1523.01 \quad (5.9)$$

The proportional coefficient β and the SN ratio η are found for the other items as well. If the SN ratio η turns out to be negative, the value will be treated as a zero in accordance with the definition given in equation (2.22) in Chapter 2, and the item will not be used in the computation of the general estimation value M . The proportional coefficients β and the SN ratio values η , found item by item, are displayed in Table 5.5.

Table 5.5. The proportional coefficients β and the SN ratios η

β, η	“B” temp	“C” temp	P 1	P 2	Pre-ht time	Manuf time
β	112.73	-968.81	-523.23	-710.78	-7.89	286.84
η	1523.01	315.26	71.21	140.46	0.00	0.00

5.1.5 COMPUTATION OF SIGNAL DATA INTEGRATED ESTIMATE VALUE

The integrated estimate yield value for Signal Data is found using the proportional coefficient β and SN ratio η (duplicate ratio), item by item. Integrated estimate yield value \hat{M}_i for the i -th Signal Data entry is found, based on the equation (5.10), by dividing value X_{ij} of the given item by incline β_j and adding it after it has been weighted with the SN ratio η_j (see Note 5.2).

$$\hat{M}_i = \frac{\eta_1 \times \frac{X_{i1}}{\beta_1} + \eta_2 \times \frac{X_{i2}}{\beta_2} + \eta_3 \times \frac{X_{i3}}{\beta_3} + \eta_4 \times \frac{X_{i4}}{\beta_4} + \eta_5 \times \frac{X_{i5}}{\beta_5} + \eta_6 \times \frac{X_{i6}}{\beta_6}}{\eta_1 + \eta_2 + \eta_3 + \eta_4 + \eta_5 + \eta_6} \quad (5.10)$$

$$= \frac{1523.01 \times \frac{X_{i1}}{112.73} + 315.26 \times \frac{X_{i2}}{(-968.81)} + 71.21 \times \frac{X_{i3}}{(-523.23)} + 140.46 \times \frac{X_{i4}}{(-710.78)} + 0 + 0}{1523.01 + 315.26 + 71.21 + 140.46 + 0 + 0}$$

Note that $X_{j1}, X_{j2}, \dots, X_{j6}$ are normalized “B” temperature, “C” temperature, Pressure 1, Pressure 2, preheating time, and manufacturing time of the i -th Signal Data. From the integrated estimate value \hat{M}_i computation equation, we can see that the larger the SN ratio η of an item, the greater the degree of its contribution to general yield estimation. Note, again, in this regard that the SN ratios η of the preheating time and manufacturing time are negative and therefore treated as zero; they will then not be used for the estimation.

Integrated estimate value \hat{M}_1 is computed below for data No. 1 of the normalized signal from Table 5.4.

$$\hat{M}_1 = \frac{1523.01 \times \frac{0.0}{112.73} + 315.26 \times \frac{44.5}{(-968.81)} + 71.21 \times \frac{39.0}{(-523.23)} + 140.46 \times \frac{46.5}{(-710.78)} + 0 + 0}{1523.01 + 315.26 + 71.21 + 140.46 + 0 + 0}$$

$$= \frac{(-28.98)}{2049.95} = -0.0141 \quad (5.11)$$

Table 5.6. Actual and integrated estimate values of yields of signal data

Data No.	Actual value M	Estimated value M
1	-3.03%	-1.41%
2	-1.59%	-1.98%
3	-1.55%	-4.72%
6	0.94%	1.43%
7	4.89%	4.67%

\hat{M}_i is found in the same way for other Signal Data as well. True value (actual value) M and the \hat{M} of the yield are shown in Table 5.6.

5.1.6 COMPUTATION OF THE SN RATIO (db) FOR INTEGRATED ESTIMATE VALUE

Procedures for computing SN ratio η (db) for integrated estimate value are shown below, using data taken from Table 5.6. The computation results will be used in the following Subsection, 5.1.7, where the evaluation of an item's relative importance will be discussed.

Linear equation

$$\begin{aligned} L &= M_1\hat{M}_1 + M_2\hat{M}_2 + \cdots + M_l\hat{M}_l \\ &= (-0.0303) \times (-0.0141) + (-0.0159) \times (-0.0198) + \cdots + 0.0489 \times 0.0467 \\ &= 0.00389 \end{aligned} \quad (5.12)$$

Effective divider

$$\begin{aligned} r &= M_1^2 + M_2^2 + \cdots + M_l^2 \\ &= (-0.0303)^2 + (-0.0159)^2 + \cdots + 0.0489^2 = 0.00389 \end{aligned} \quad (5.13)$$

Total variation

$$\begin{aligned} S_T &= \hat{M}_1^2 + \hat{M}_2^2 + \cdots + \hat{M}_l^2 \\ &= (-0.0141)^2 + (-0.0198)^2 + \cdots + 0.0467^2 = 0.00520 \quad (f=5) \end{aligned} \quad (5.14)$$

Variation S_β of a proportional term

$$S_\beta = \frac{L^2}{r} = \frac{0.00389^2}{0.00389} = 0.00389 \quad (f=1) \quad (5.15)$$

Error variation

$$S_e = S_T - S_\beta = 0.00520 - 0.00389 = 0.00131 \quad (f=4) \quad (5.16)$$

Error variance

$$V_e = \frac{S_e}{l-1} = \frac{0.00131}{4} = 0.00033 \quad (5.17)$$

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

$$\begin{aligned}\eta &= 10 \log \left(\frac{\frac{1}{r} (S_\beta - V_e)}{V_e} \right) = 10 \log \left(\frac{\frac{1}{0.00389} (0.00389 - 0.00033)}{0.00033} \right) \\ &= 10 \log (2795.97853) = 34.47 \text{ (db)}\end{aligned}\quad (5.18)$$

5.1.7 EVALUATION OF THE IMPORTANCE OF AN ITEM

Thus far, we have been performing general estimation using all six items. But the six items include some that are effective for integrated estimation purposes as well as others that are not very effective. Therefore, item importance evaluation is performed with the use of an orthogonal array.

5.1.7.1 Item Layout for Orthogonal Array L_{12}

We have six items to deal with here, so an L_{12} type orthogonal array is chosen for item allotment. The reason for choosing an orthogonal array L_{12} , which is a $4 \times$ prime-type array, is that it will enable the minimization of the impact of any interaction between the items, should any occur (see Note 5.3).

5.1.7.2 Computation of the SN Ratio (db) of the Orthogonal Array by Row

The six items, from “B” temperature to manufacturing time, are allotted to Columns 1 through 6 (Table 5.7). Numerals “1” and “2” in the Table indicate the level at which a given item is allotted to the column, and thus:

Level 1: The item will be used.

Level 2: The item will *not* be used.

In Test No. 1, the data in Columns 1 through 6 are all at Level 1, which signifies that all six items will be used. For the SN ratio of integrated estimation, the SN ratio of η (= 34.47 db), which is found in the previous Subsection, 5.1.6, will be used.

In Test No. 2, because Columns 1 to 5 are at Level 1, five items, “B” temperature through pre-heating time are used; and it is shown that the integrated estimate SN ratio is 34.47 (db). Likewise, in Test No. 12, because Columns 3, 4, and 6 are at Level 1, three items, Pressure 1, Pressure 2, and manufacturing time, are used, and it is shown that the SN ratio is 20.65 (db).

Table 5.7. Orthogonal array L_{12} and the layout of items

No.	“B” temp	“C” temp	P 1	P 2	Pre-heat time	Manuf time	e	e	e	e	e	Integrated estimate SN ratio (db)
1	1	1	1	1	1	1	1	1	1	1	1	34.47
2	1	1	1	1	1	2	2	2	2	2	2	34.47
3	1	1	2	2	2	1	1	1	2	2	2	33.87
4	1	2	1	2	2	1	2	2	1	1	2	32.64
5	1	2	2	1	2	2	1	2	1	2	1	33.16
6	1	2	2	2	1	2	2	1	2	1	1	31.83
7	2	1	2	2	1	1	2	2	1	2	1	24.99
8	2	1	2	1	2	2	2	1	1	1	2	24.16
9	2	1	1	2	2	2	1	2	2	1	1	24.29
10	2	2	2	1	1	1	1	2	2	1	2	21.48
11	2	2	1	2	1	2	1	1	1	2	2	18.53
12	2	2	1	1	2	1	2	1	2	2	1	20.65

(Note: Level 1: “Item will be used.” Level 2: “Item will *not* be used.” Here, items treated as SN ratio $\eta=0$ in Table 5.5 are allotted as well.)

5.1.7.3 Creation of an Integrated Estimate SN Ratio (db) Factorial Effect Graph

On the basis of Table 5.7, an auxiliary table to the integrated estimate SN ratio (db) is sought and the resulting data is displayed in Table 5.8.

Table 5.8. Integrated estimate SN ratio (db) auxiliary table (averages by level)

Item	Level 1	Level 2
“B” temperature	33.41	22.35
“C” temperature	29.37	26.38
Pressure 1	27.51	28.25
Pressure 2	28.06	27.69
Preheating time	27.62	28.13
Manufacturing time	28.02	27.74

Going further, based on this auxiliary table, Figure 5.3, a factorial effect graph is created. From this factorial effect graph we learn how these “B” and “C” temperatures, with inclines dropping steeply from Level 1 to Level 2, play significant roles in yield prediction.

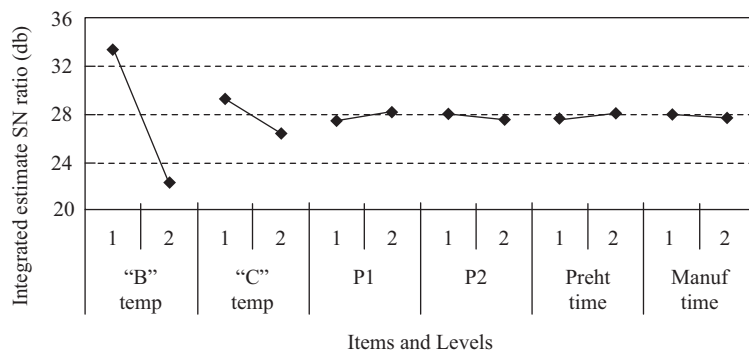


Figure 5.3. Factorial effects for the six items.

5.1.7.4 Computation of Integrated Estimate Value \hat{M} for Signal Data Under Optimum Conditions

A comparison between the SN ratio using only “B” and “C” temperatures, which are assumed to be significant contributors to yield prediction based on the factorial effect graph, versus all of the items is shown in Table 5.9. It is understood that there is not a substantial difference between them.

Table 5.9. Comparison of integrated estimate SN ratios (db)

Case	Items used	Integrated estimate SN ratio (db)
1	All items	34.47
2	“B” temperature, “C” temperature	33.87

5.1.8 INTEGRATED YIELD ESTIMATION FOR UNKNOWN DATA

From the subsequent manufacturing/processing we have acquired some unknown data, as shown in Table 5.10. On the basis of the data, the yield is submitted to integrated estimation.

For the unknown data in Table 5.11, as well, normalization is executed by subtracting from it the average value of each of the items in the Unit Space. Normalized unknown data is shown in Table 5.11.

Table 5.10. Unknown data

Data No.	“B” temp	“C” temp	P 1	P 2	Pre-ht time	Manuf time	Yield
<1>	563.0	306.5	185.5	183.5	2.8	60.0	unknown

Table 5.11. Normalized unknown data

Data No.	“B” temp	“C” temp	P 1	P 2	Pre-ht time	Manuf time	Yield
<1>	-12.0	77.0	19.0	19.5	-4.2	-60.0	unknown

With normalized unknown data <1>, as in other cases we have described, integrated estimate yield value $\hat{M}_{<1>}$ is computed using equation (5.10).

$$\begin{aligned}\hat{M}_{<1>} &= \frac{1523.01 \times \frac{(-12.0)}{112.73} + 315.26 \times \frac{77.0}{(-968.81)} + 71.21 \times \frac{19.0}{(-523.23)} + 140.46 \times \frac{19.5}{(-710.78)} + 0 + 0}{1523.01 + 315.26 + 71.21 + 140.46 + 0 + 0} \\ &= \frac{-193.62}{2049.95} = -0.0945\end{aligned}\quad (5.19)$$

5.1.9 COMPUTATION OF INTEGRATED ESTIMATE VALUES BEFORE NORMALIZATION

Through the computations performed so far, yields (output values) have been normalized in terms of the average Unit Space values. Therefore, in order to find the integrated estimate value \hat{y} of the actual yield, the average value M_0 (=84.58%) of the yield of the Unit Space is added to the normalized value \hat{M} . For instance, the integrated estimate value \hat{y}_1 for the yield of Signal Data No. 1 is found as follows:

$$\hat{y}_1 = \hat{M}_1 + M_0 = -1.41 + 84.58 = 83.17 \text{ (\%)} \quad (5.20)$$

For the unknown data as well, the integrated estimate value \hat{y} of the yield is found by adding the yield average value M_0 (= 84.58) of the Unit Space to the likewise normalized value \hat{M} . Therefore, the yield's integrated estimate value $\hat{y}_{<1>}$ for unknown data <1> will be:

$$\hat{y}_{<1>} = \hat{M}_{<1>} + M_0 = -9.45 + 84.58 = 75.13 \text{ (\%)} \quad (5.21)$$

Subsequently we have learned that the actual yield value $y_{<1>}$ acquired for unknown data is 73.30%. The actual value y and the generally estimated value \hat{y} of the yields for Signal Data and unknown data are shown in Table 5.12.

Table 5.12. The Actual value y and the integrated estimate value \hat{y} of the yields for signal data and unknown data

No.		Actual value y	Integrated estimate value \hat{y}
Signal Data	1	81.55%	83.17%
	2	82.99%	82.60%
	3	83.03%	79.86%
	6	85.52%	86.01%
	7	89.47%	89.25%
Unknown data	<1>	73.30%	75.13%

Figure 5.4 shows a scatter diagram of the actual values and integrated estimate values of the yields for the five Signal Data entries and one piece of unknown data listed in Table 5.12. The horizontal axis represents the actual values of the yield, y , and the vertical axis the integrated estimate values of the yield, \hat{y} .

We learn from the figure that the plotted dots overlap rather closely, almost a 45° straight line, indicating that the Signal Data as well as the unknown data are fairly good estimations.

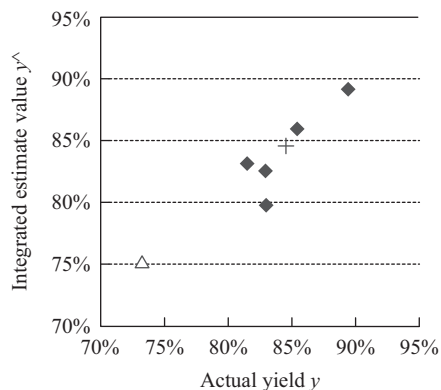


Figure 5.4. Actual yield value y and integrated estimate value \hat{y} .

Note 5.1 Omega transformation (Logit transformation)

Dr. Taguchi's work in tile quality improvement assistance through applied orthogonal array tests in the 1950s is today legendary. In the context of the same work, he once encountered a case in which he added up the improvement effects upon the percent defective, which was attributable to various factors, only to find to his surprise that the percent defective figure had come out to a negative value. Realizing that percentage figures could not be added up, he proposed and used a transformation formula that would make it possible to perform percentage figure addition and named it "omega [Ω] transformation" after Fischer's Z conversion of correlative coefficients. Omega transformation is the same thing as the logit transformation used in logistic regression in statistics. This is also the same clarity conversion formula used in telephony, as defined in 1952 by the CCIF (International Telephone Advisory Committee). Dr. Taguchi remarked:

"In a situation where data is accepted only within the limits of 0 to 1 (or 0% to 100%), there is an alternative way of finding the SN ratio by first performing an omega transformation (rather than by using raw data). ... Omega transformation is a method of converting the variable domain from (0, 1) to $(-\infty, \infty)$ through processing, using the following equation in which all ratios take a value within the parameters of $0 < p < 1$, as follows:

$$y = -10 \log \left(\frac{1}{p} - 1 \right) \quad (5.22)$$

"The purpose of the transformation is to allow arithmetical addition to be performed of the effects of various causes with impact on percentage rate data."

In other words, where, as with yield, the permitted scope is restricted to 0% to 100%, it is advisable for the data first to be omega-transformed before being submitted to analysis.⁶ The relationship between yield and omega transformation is illustrated in Figure 5.5.

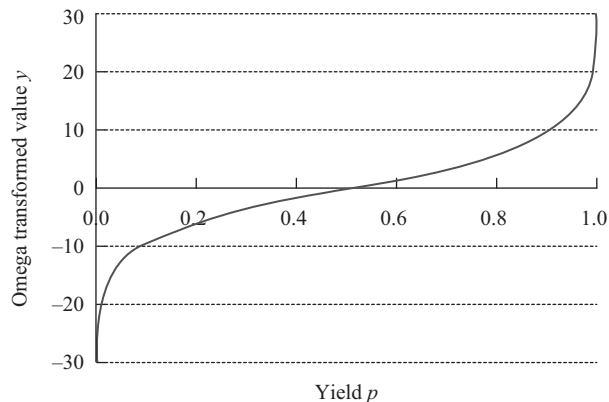


Figure 5.5. Relationship of yield p to omega-transformed value y .

Note 5.1 Omega transformation (Logit transformation) (Cont.)

If addability is to be established as a valid factor, all percentage values, strictly speaking, should be omega-transformed, but if operational transformations are an issue, a generally followed practice is to omit the conversion altogether if the percentage value in question is within 30–70%; transformation is a sine qua non if less than 10% or more than 90% of data is included. Appropriate judgment on a case-by-case basis is necessary if the data is within 20–80% (see Figure 5.5).

Kodak has defined the photographic and reprographic image density scale with Macbeth density $D = \log 1/p$ (p : permeability or reflectance ratio), but the given formula is p [white 1, black 0] $\rightarrow D$ [white 0, black $+\infty$] with no room for addability. For this reason, the Taguchi method uses the omega-converted value [white $-\infty$, black $+\infty$] of the absorption coefficient $q (= 1 - p)$ [white 0, black 1].

Dr. Yoshiko Yokoyama, addressing this issue, observed: “Think of an eigenvalue called *percent defective*. Suppose that the percent defective of a process is 2%. It would be quite a challenge to bring it down into the vicinity of 1%. But if the percent defective stands at, say, 30%, it will not be so difficult to reduce it by 1%. What this means is, it can be said, that an entirely different sort of difficulty is involved when reduction of 2% to 1% is the issue and when reduction of 30% to 29% is the issue, even though both cases are about a reduction of 1%.”

In dealing with properties with a 0–100% constraint, then, one would do well to first submit the data to omega transformation before performing the evaluation.

We now examine what relevance omega transformation may have with respect to yield prediction for a manufacturing process, which is the theme of Section 5.1. When omega transformation is applied to the yield of 89.47% for Signal Data 7 in Table 5.12 and the yield of 73.30% for unknown data <1>, we obtain the results displayed in Table 5.13.

Table 5.13. Omega-transformed yield values

	Yield p	Omega-Transformed Value y
(a) Yield of 89.47% Improved by 1%		
After improvement	0.9047	9.77
Current	0.8947	9.29
Difference	0.0100	0.48
(b) Yield of 73.30% Improved by 1%		
After improvement	0.7430	4.61
Current	0.7330	4.39
Difference	0.0100	0.22

The omega transformation makes it clear that improving an 89.47% yield by 1% requires more than two times as much improvement efforts as improving a 73.30% yield by 1%.

Note 5.2 Treatment of η and β in integrated estimate equations

Some readers may wonder: “Whereas the integrated estimate equation has η_j and β_j of all items left intact, why is it that they are not expressed as coefficients item by item in the form of $\hat{M} = a_1x_{i1} + a_2x_{i2} + \dots + a_kx_{ik}$, as would be the case in the regression equation?”

The integrated estimate equation is expressed as follows, as explained in Subsection 2.5.4:

$$\hat{M}_i = \frac{\eta_1 \times \frac{X_{i1}}{\beta_1} + \eta_2 \times \frac{X_{i2}}{\beta_2} + \dots + \eta_k \times \frac{X_{ik}}{\beta_k}}{\eta_1 + \eta_2 + \dots + \eta_k} \quad (i = 1, 2, \dots, l) \quad (5.23)$$

Here, given that $(\hat{M}_{ij} =) X_{ij}/\beta_j$ is the estimated output value \hat{M}_{ij} found from zero point proportional equation $X_{ij} = \beta_j M_{ij}$ with respect to item j , equation (5.23) can also be expressed as follows:

$$\hat{M}_i = \frac{\eta_1 \times \hat{M}_{i1} + \eta_2 \times \hat{M}_{i2} + \dots + \eta_k \times \hat{M}_{ik}}{\eta_1 + \eta_2 + \dots + \eta_k} \quad (i = 1, 2, \dots, l) \quad (5.24)$$

From equation (5.24), we see that the integrated estimate value, \hat{M}_i , is arrived at by applying estimated value \hat{M}_{i2} for item 2, ..., and estimated value \hat{M}_{ik} for item k for weighted averages, $\eta_1, \eta_2, \dots, \eta_k$, which indicate the relative degrees of importance of the items, to estimated value \hat{M}_{i1} for item 1.

Furthermore, integrated estimate equation (5.23) can also be developed as follows:

$$\begin{aligned} \hat{M}_i &= \frac{\eta_1 \times \frac{X_{i1}}{\beta_1} + \eta_2 \times \frac{X_{i2}}{\beta_2} + \dots + \eta_k \times \frac{X_{ik}}{\beta_k}}{\eta_1 + \eta_2 + \dots + \eta_k} \\ &= \left(\frac{\eta_1}{\beta_1 \sum_{j=1}^k \eta_j} \right) \times X_{i1} + \left(\frac{\eta_2}{\beta_2 \sum_{j=1}^k \eta_j} \right) \times X_{i2} + \dots + \left(\frac{\eta_k}{\beta_k \sum_{j=1}^k \eta_j} \right) \times X_{ik} \quad (i = 1, 2, \dots, l) \\ &= a_1 X_{i1} + a_2 X_{i2} + \dots + a_k X_{ik} \end{aligned} \quad (5.25)$$

In view of the above, we understand that the general estimation formula can be expressed as follows, as in the case of multiple regression analysis:

$$\hat{M} = a_1 X_{i1} + a_2 X_{i2} + \dots + a_k X_{ik} \quad (5.26)$$

Note 5.2 Treatment of η and β in integrated estimate equation (Cont.)

If so, why is it that instead of the integrated estimate equation being expressed in the same way the regression equation is, the η_j and β_j of each of its items are left intact? The answer is: this is for the sake of making it easier to see in detail how the degree of impact of each item is taken into consideration by weighting the estimated value of each item with η_j . Another reason is that the digits that follow the decimal point of coefficient a_i will assume increasingly smaller values such that rounding off will produce effect-distorting errors unless many significant digits are assigned.

Note 5.3 What is a $4 \times$ prime two-level orthogonal array?

“The importance of a given item,” says Dr. Taguchi, “is evaluated in terms of the degree to which deterioration occurs when the item in question is done without, not in its SN ratio, but in the SN ratio η of integrated estimation. It is thus advisable to use a two-level orthogonal array (even more advisably, a $4 \times$ prime two-level type, specifically). ... The purpose of using an orthogonal array for making comparisons of SN ratios η (in the context of integrated estimation) is that, in finding out to what degree each of the items is relevant toward maximizing the accuracy of prediction and estimation. If the evaluation is to have any reliability, each such item needs to be comparatively evaluated under various conditions.”

In short, we recommend the use of a two-level-type orthogonal array in the evaluation of the relative importance of items.

A “two-level orthogonal array with a $4 \times$ prime number” refers to any of the following series:

$$L_{12} (4 \times 3), L_{20} (4 \times 5), L_{28} (4 \times 7), L_{44} (4 \times 11), L_{52} (4 \times 13), L_{68} (4 \times 17), \text{ etc.}$$

The cyclic orthogonal array of the $4 \times$ prime number type created using Paley’s method, in particular, is known for the almost completely even scatter of interactions between two columns. For more on how to create these orthogonal arrays, refer to Chapter 7.

5.2 CHARACTER PATTERN RECOGNITION USING THE RT METHOD

With the RT Method, no matter how great the number of items is, they are condensed into two variables. In this Section, we use patterns drawn on boards with 5×7 grids, each featuring a numeral character pattern consisting of a total of 35 picture elements. In Chapter 4, Section 4.1, we did some exercises in character pattern characterization. For the purposes of the work in this Section, however, we define the white grid as zero (0), and the black grid as one (1). That means that we will

have, per character, a set of 35 pieces of data, each either 0 or 1. We will perform character recognition based on the two variables into which the 35 items have been consolidated.

5.2.1 DEFINITION OF THE UNIT SPACE

The Unit Space here is defined as a *group of characters readable as the numeral “7”*; and 16 such character patterns have been prepared. Of the 16 character patterns, character patterns Nos.1 and 15 are displayed below in Figure 5.6.

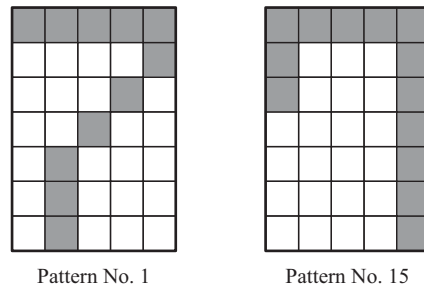


Figure 5.6 Sample patterns of “7” unit space.

The 16 Unit Space characters readable as “7” are each represented, in Table 5.14, as a set of 35 pieces of data, each as either 0 or 1. In the table below, the samples appear in the vertical direction, and the items in the horizontal direction.

“Average value” refers to the Unit Space average, the items of which have all been accounted for. Item No. 35, for example, examined in greater detail here, is as follows:

$$\bar{x}_{35} = \frac{1}{16}(0 + \cdots + 0 + 1 + 1 + 1 + 0) = \frac{3}{16} = 0.1875 \quad (5.27)$$

Note that, due to lack of space, the average values in Table 5.14 are rounded off at two digits after the decimal point.

As for linear equation L and effective divider r , they are found from equations (5.29) and (5.30) in Subsection 5.2.2, and retain digits as far down from the decimal point as the fourth digit. As demonstrated in Chapter 2, Table 2.13, linear equation L is displayed sample by sample in the (L, r) column, and effective divider r is displayed in the bottom row for average values in the (L, r) column because, in the case of the Unit Space data, the values displayed are average values for linear equations L_1, L_2, \dots, L_n .

Data

Teshima, S., Hasegawa, Y., & Tatebayashi, K. (2012). Quality recognition & prediction : Smarter pattern technology with the mahalanobis-taguchi system. ProQuest Ebook Central http://ebookcentral.proquest.com/created from britishcouncilonline-ebooks on 2020-05-26 17:05:42

5.2.2 COMPUTATION OF SENSITIVITY β AND STANDARD SN RATIO OF UNIT SPACE SAMPLES

Using the RT Method computation formula discussed in Subsection 2.5.6, we will find sensitivity β and standard SN ratio η (in the form of a duplicate ratio) for each Unit Space sample. How to find sensitivity β and standard SN ratio η (in the form of a duplicate ratio) for Unit Space sample No. 1 is explained here:

$$\text{Sensitivity } \beta_1 = \frac{L_1}{r} = \frac{7.6875}{8.9141} = 0.8624 \quad (5.28)$$

Where for the linear equation, we have:

$$L_1 = 1 \times 1 + 1 \times 1 + \dots + 0 \times 0 + 0.1875 \times 0 = 7.6875 \quad (5.29)$$

and, for the effective divider,

$$r = 1^2 + 1^2 + \dots + 0^2 + 0.1875^2 = 8.9141 \quad (5.30)$$

Standard SN ratio η_1 (duplicate ratio) is computed after all fluctuations of S_{T1} , and the proportional item variation $S\beta_1$, etc. have first been found.

Total variation

$$S_{T1} = 1^2 + 1^2 + \dots + 0^2 + 0^2 = 11 \quad (f=35) \quad (5.31)$$

Variation of proportional term $S\beta_1$

$$S_{\beta 1} = \frac{L_1^2}{r} = \frac{7.6875^2}{8.9141} = 6.6297 \quad (f=1) \quad (5.32)$$

Error variation

$$S_{e1} = S_{T1} - S_{\beta 1} = 11 - 6.6297 = 4.3703 \quad (f=34) \quad (5.33)$$

Error variance

$$V_{e1} = \frac{S_{e1}}{k1} = \frac{4.3703}{34} = 0.1285 \quad (5.34)$$

Accordingly, standard SN ratio η_1 (duplicate ratio) is given as the following equation:

$$\eta_1 = \frac{1}{V_{e1}} = \frac{1}{0.1285} = 7.7798 \quad (5.35)$$

In the same manner, sensitivity β and standard SN ratio η (duplicate ratio) have been found for each individual sample of the Unit Space separately and are tabulated in Table 5.15.

Table 5.15. Sensitivity β and standard SN ratio η (duplicate ratio) for each individual unit space sample

Sample No.	Sensitivity β	SN ratio η
1	0.8624	7.7798
2	0.9465	8.4712
3	1.0026	8.4179
4	0.9816	7.7079
5	0.9606	12.2514
6	1.0447	14.9697
7	1.1008	15.4650
8	1.0657	11.8238
9	0.9886	14.8608
10	1.0727	19.5190
11	1.1288	20.7178
12	1.0657	11.8238
13	0.8484	7.4169
14	0.9325	8.0029
15	0.9886	7.9293
16	1.0096	8.6885

5.2.3 COMPUTATION OF TWO VARIABLES Y_1 AND Y_2 , UNIT SPACE SAMPLE BY SAMPLE

We compute the two variables Y_1 and Y_2 using sensitivity β and SN ratio η (duplicate ratio) in Table 5.15. The new variable Y_1 will retain sensitivity β in intact form, but Y_2 will not retain the intact form of SN ratio η , which will be converted to enable the evaluation of the state of scatter from the standard condition (average value of the Unit Space character “7” in Table 5.14). The computation of Y_1 and Y_2 for Unit Space sample No. 1 in Table 5.15 works out as follows:

$$Y_1 = \beta_1 = 0.8624 \quad (5.36)$$

$$Y_2 = \frac{1}{\sqrt{\eta_1}} = \frac{1}{\sqrt{7.7798}} = 0.3585 \quad (5.37)$$

In the same way, we compute Y_1 and Y_2 for all the Unit Space samples; the average values for both as found are shown in Table 5.16.

Table 5.16. Y_1 and Y_2 for the unit space samples

Sample No.	Y_1	Y_2
1	0.8624	0.3585
2	0.9465	0.3436
3	1.0026	0.3447
4	0.9816	0.3602
5	0.9606	0.2857
6	1.0447	0.2585
7	1.1008	0.2543
8	1.0657	0.2908
9	0.9886	0.2594
10	1.0727	0.2263
11	1.1288	0.2197
12	1.0657	0.2908
13	0.8484	0.3672
14	0.9325	0.3535
15	0.9886	0.3551
16	1.0096	0.3393
Average	1.0000	0.3067

5.2.4 COMPUTATION OF DISTANCES OF UNIT SPACE SAMPLES

After going through the preceding procedures, we have before us two variables, Y_1 and Y_2 , which consolidate the 35 feature values of each of the Unit space samples. Next, using the MT Method, the Mahalanobis Distance, MD , of each of the Unit Space samples from the center of the Unit Space is found. The result of this procedure is displayed in Table 5.17.

For a discussion of computations revolving around the MT Method, refer to Subsections 2.3.2 and 2.5.1. If the RT Method is chosen, the computation process will be identical to that which is discussed in 2.3.2, since computation of two-dimensional Mahalanobis Distances will be involved. The correlation coefficient between Y_1 and Y_2 in this case will be -0.779 . Note also that the average value of the MD for Unit Space data using the MT Method is 1.0.

Table 5.17. Distances of unit space samples

Sample No.	MD(D ²)
1	1.770
2	0.291
3	0.788
4	1.023
5	0.990
6	0.504
7	0.863
8	0.521
9	1.449
10	1.415
11	1.650
12	0.521
13	2.087
14	0.467
15	0.938
16	0.722
Average	1.000

5.2.5 SIGNAL DATA

In connection with performing an evaluation of discrimination ability, we prepared as Signal Data two groups of 12 characters, one a group of characters readable as the numeral “1,” and another a group of characters readable as the numeral “9,” which resemble the Unit Space character “7.” If discrimination is possible with characters that are similar, it will be easy to discriminate dissimilar characters. Of the 12 character patterns in the Signal Data set of characters readable as “1,” character patterns Nos. 1 and 4 are shown in Figure 5.7.

As for the 12 patterns for the Signal Data “1,” the 35 data sets (0, 1) are as shown in Table 5.18. The samples appear in the vertical direction, and the items in the horizontal direction.

As for Signal Data “9,” out of the 12 character patterns prepared for it, patterns Nos. 1 and 5 are shown below as examples in Figure 5.8.

In the same manner as described earlier, the 35 data sets (0, 1) for the 12 patterns for Signal Data for “9,” are shown in Table 5.19. The samples appear in the vertical direction, and the items in the horizontal direction.

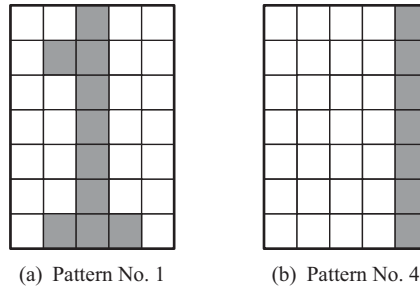


Figure 5.7. Sample signal data “1” patterns.

5.2.6 Sample-by-Sample Computation of Sensitivity and Standard SN Ratio for Signal Data

As is the case for Unit Space, sensitivity β and standard SR ratio η (duplicate ratio) are found for each individual sample of the signal. We will explain below how to find sensitivity β_1 and standard SN ratio η_1 (duplicate ratio) for sample No. 1 of signal “1.”

$$\text{Sensitivity } \beta_1 = \frac{L'_1}{r} = \frac{2.6875}{8.9141} = 0.3015 \quad (5.38)$$

where:

Linear equation

$$L'_1 = 1 \times 0 + 1 \times 0 + \cdots + 0 \times 1 + 0.1875 \times 0 = 2.6875 \quad (5.39)$$

Total variation

$$S_{T1} = 0^2 + 0^2 + \cdots + 1^2 + 0^2 = 10 \quad (f=35) \quad (5.40)$$

Variation of proportional term $S_{\beta 1}$

$$S_{\beta 1} = \frac{L'^2_1}{r} = \frac{2.6875^2}{8.9141} = 0.8103 \quad (f=1) \quad (5.41)$$

Error variation

$$S_{e1} = S_{T1} - S_{\beta 1} = 10 - 0.8103 = 9.1897 \quad (f=34) \quad (5.42)$$

Table 5.18. Data and linear equations for signal character “1”

Data																																				
No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	L'
1	0	0	1	0	0	0	1	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	1	1	0	2.6875		
2	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	2.2500		
3	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	1	0	2.2500			
4	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	3.5625	
5	1	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	3.1875		
6	0	0	1	0	0	0	1	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	2.2500		
7	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	1	1	1	1	1	3.0625	
8	0	0	1	0	0	1	1	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	1	1	1	1	1	1	3.8125	
9	0	0	1	0	0	0	1	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	1	1	1	0	2.6875	
10	0	0	1	0	0	0	1	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	1	1	1	0	2.6875	
11	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	1	1	1	0	2.6875	
12	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	1	1	1	1	1	3.0625	

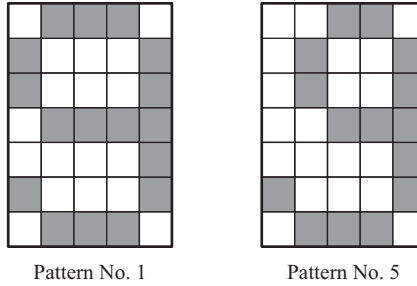


Figure 5.8. Sample patterns for signal data “9”.

Error variance

$$V_{el} = \frac{S_e}{k-1} = \frac{9.1897}{34} = 0.2703 \quad (5.43)$$

From this, we move to finding standard SN ratio η_l (duplicate ratio) from the following equation:

$$\eta_1 = \frac{1}{V_{el}} = \frac{1}{0.2703} = 3.6998 \quad (5.44)$$

Likewise, sensitivity β and standard SN ratio η (duplicate ratio) are found for each instance of Signal Data “1” and Signal Data “9.” They are shown in Table 5.20.

5.2.7 COMPUTATION OF TWO VARIABLES, Y_1 AND Y_2 , FOR EACH SIGNAL SAMPLE

We now compute the two variables, Y_1 and Y_2 , using the two items, sensitivity β and standard SR ratio η (duplicate ratio), from Table 5.20. Here, again, standard SN ratio η is not used as-is but first undergoes transformation to enable the evaluation of the state of scatter from the Unit Space. Based on the data in Table 5.20, Y_1 and Y_2 are computed for sample No. 1 of the signal “1,” the process working out as follows:

$$Y_1 = \beta_1 = 0.3015 \quad (5.45)$$

$$Y_2 = \frac{1}{\sqrt{\eta_1}} = \frac{1}{\sqrt{3.6998}} = 0.5199 \quad (5.46)$$

Table 5.19. Data and linear equation for signal character “g”

Data																																				
No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	L, r
1	0	1	1	1	0	1	0	0	0	1	1	0	0	0	1	0	1	1	1	1	0	0	0	0	1	1	0	0	0	1	0	1	1	1	0	8.0000
2	0	1	1	1	0	1	0	0	1	1	0	1	0	0	1	0	1	1	1	1	0	0	0	0	1	0	0	0	0	1	1	1	1	1	0	8.1875
3	1	1	1	1	1	0	0	1	1	0	0	1	0	0	1	0	1	1	1	1	0	0	0	1	1	0	0	1	1	0	1	1	1	1	1	10.3750
4	1	1	1	1	1	0	0	1	1	0	0	1	0	0	1	0	1	1	1	1	0	0	0	0	1	0	0	0	0	1	1	1	1	1	1	10.3750
5	0	0	1	1	0	1	0	1	0	1	0	1	0	0	1	0	1	0	1	1	0	0	0	1	1	0	0	0	0	1	0	1	1	1	0	5.7500
6	0	0	1	1	0	1	0	1	0	1	0	1	0	0	1	0	0	1	1	1	0	0	0	0	1	0	0	0	0	1	1	1	1	1	0	5.9375
7	0	0	1	1	1	0	1	0	1	0	1	0	0	1	0	0	1	0	1	1	0	0	0	0	1	1	0	0	0	1	1	1	1	1	1	7.1250
8	0	0	1	1	1	0	1	0	1	0	1	0	0	1	0	1	0	1	1	1	0	0	0	0	1	0	0	0	0	1	1	1	1	1	1	7.1250
9	0	0	1	1	1	0	1	0	1	0	1	0	0	1	0	1	0	1	1	1	0	0	0	0	1	1	0	0	0	1	0	1	1	1	0	6.7500
10	0	0	1	1	0	0	1	0	1	0	1	0	0	1	0	0	1	0	1	1	0	0	0	0	1	1	0	0	0	1	1	1	1	1	0	5.9375
11	0	1	1	0	0	1	0	0	1	0	1	0	0	0	1	0	1	1	1	1	0	0	0	0	1	1	0	0	0	1	1	1	1	1	1	6.3750
12	0	1	1	0	0	1	0	0	1	0	1	0	0	0	1	0	1	1	1	1	0	0	0	0	1	1	0	0	0	1	0	0	1	1	1	6.1875

Table 5.20. Sensitivity β and standard SR ratio η (duplicate ratio) for each instance of signal data

Sample No.	Character 1		Character 9	
	Sensitivity β	SN ratio η	Sensitivity β	SN ratio η
1	0.3015	3.6998	0.8975	3.4622
2	0.2524	4.5748	0.9185	3.5866
3	0.2524	5.2860	1.1639	3.8097
4	0.3996	6.0973	1.1639	4.2904
5	0.3576	4.9561	0.6450	3.0113
6	0.2524	4.5748	0.6661	3.0783
7	0.3436	3.1056	0.7993	2.7631
8	0.4277	2.9905	0.7993	3.0075
9	0.3015	3.6998	0.7572	2.8599
10	0.3015	3.6998	0.6661	2.8227
11	0.3015	4.1515	0.7152	2.5296
12	0.3436	3.4178	0.6941	2.6761

In the same way, Y_1 and Y_2 are computed sample by sample. The outcome is tabulated in Table 5.21.

Table 5.21. Y_1 and Y_2 computed for each sample

Sample No.	Character 1		Character 9	
	Y_1	Y_2	Y_1	Y_2
1	0.3015	0.5199	0.8975	0.5374
2	0.2524	0.4675	0.9185	0.5280
3	0.2524	0.4349	1.1639	0.5123
4	0.3996	0.4050	1.1639	0.4828
5	0.3576	0.4492	0.6450	0.5763
6	0.2524	0.4675	0.6661	0.5700
7	0.3436	0.5674	0.7993	0.6016
8	0.4277	0.5783	0.7993	0.5766

(Continued on following page)

Table 5.21. (Continued)

Sample No.	Character 1		Character 9	
	Y_1	Y_2	Y_1	Y_2
9	0.3015	0.5199	0.7572	0.5913
10	0.3015	0.5199	0.6661	0.5952
11	0.3015	0.4908	0.7152	0.6287
12	0.3436	0.5409	0.6941	0.6113

We thereby arrive at the condensation, through all the above computations, of the values of all Signal Data into the two variables Y_1 and Y_2 . Working on the basis of Table 5.16 and Table 5.21, we create a scatter diagram (Figure 5.9) for Y_1 and Y_2 . A look at this scatter diagram makes it plain to see how Unit Space “7” forms an aggregation of its own while the Signal Data for “1” as well as the Signal Data for “9” ends up scattered at distinct distances.

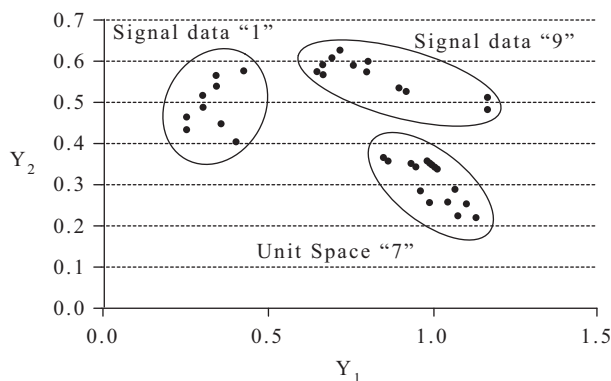


Figure 5.9. Y_1 – Y_2 scatter diagram for unit space and signal data.

5.2.8 MAHALANOBIS DISTANCE COMPUTED FOR EACH SAMPLE

Mahalanobis Distance D is found for each of the samples of the signal using the MT Method. The resulting solution is shown in Table 5.22. We see how great the Mahalanobis Distance turns out for both Signal Data “1” and Signal Data “9,” proving that they are not of the same pattern as “7.”

Table 5.22. Mahalanobis Distance (D^2) measurements of the signal, sample by sample

Data No.	Character “1”	Character “9”
1	51.51	17.19
2	71.66	17.10
3	79.45	44.79
4	52.22	36.53
5	52.11	14.80
6	71.66	13.97
7	39.20	22.46
8	27.97	17.85
9	51.51	18.32
10	51.51	16.72
11	56.12	22.99
12	41.51	19.36

5.2.9 COMPUTATION OF MAHALANOBIS DISTANCE FOR UNKNOWN DATA

We created five new characters readable as “7” as unknown data. Of the five character patterns, character patterns Nos. 2 and 4 are shown in Figure 5.10.

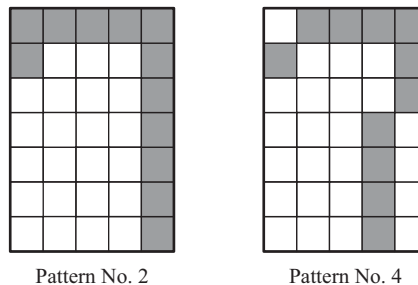
**Figure 5.10.** Sample patterns for unknown data “7”.

Table 5.23 displays the 35 (0,1) data sets for the five unknown pieces of data. The samples appear in the vertical direction, and items in the horizontal direction.

As in the case of Signal Data, with unknown data as well, sensitivity β and standard SR ratio η are computed using the average value of each item in the Unit Space. Then, using the sensitivity β and standard SR ratio η of the unknown data, the two variables Y_1 and Y_2 are found. Lastly, the distance, D , from the center of the Unit Space is computed. The result is displayed in Table 5.24.

All patterns are found to yield modest MD readings. The generally accepted threshold for the MD is of the order of 4, and so these readings in the table can be judged to be of the same pattern as “7.”

Table 5.24. Mahalanobis distance (D^2) readings for unknown data

Data No.	MD(D^2)
1	1.628
2	1.765
3	3.055
4	0.922
5	2.900

Note 5.4 Utilizing the Multi RT Method when dealing with multiple items

The MT System family includes a method called the “Multi Method” (see Chapter 2, Subsection 2.1.5). The Multi Method proves useful when used to deal with cases featuring particularly large numbers of items. This method has not been discussed in a dedicated manner so far in this book, and a subject-specific, though concise, explanation is offered here of the underlying concept of the Multi RT Method using an example of its application to pictorial image recognition.

The method of choice for identifying a large number of items by condensing them into the two variables of Y_1 and Y_2 is the RT Method. Let’s suppose here that we have n number of pieces of image data each comprising a total of $500 \times 500 = 250,000$ picture elements (see Figure 5.11), and that this body of data is homogeneous (as of a certain character, or a certain animal; or if the data were an x-ray photograph revealing no evidence of illness).

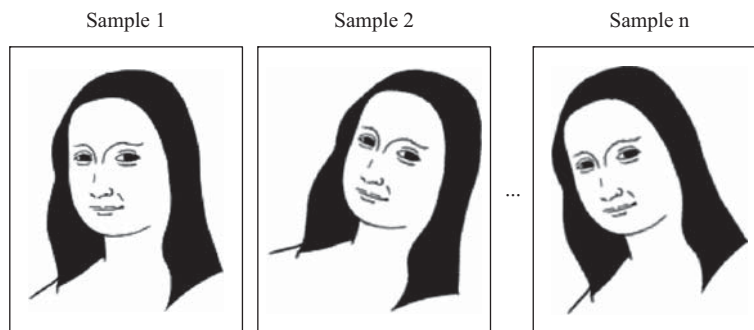


Figure 5.11. Unit space members of 500×500 elements.

Images generally involve enormous quantities of data. In cases such as we have here, if, say, the brightness of each picture element is to be treated as an item, we will be dealing with 250,000 items. Unless something is done about it, this may make computations difficult to accomplish due to memory size and other constraints. One conceivable solution to such a problem would be, for example, as shown in Figure 5.12, to divide the given image into $50 \times 50 = 2,500$ picture elements to create 100 partial images before re-synthesizing them, which is the strategy employed by the Multi RT Method. For this method, the 100 partial images will be labeled G_1, G_2, \dots, G_{100} in a Multi RT Method procedure, which is explained below.

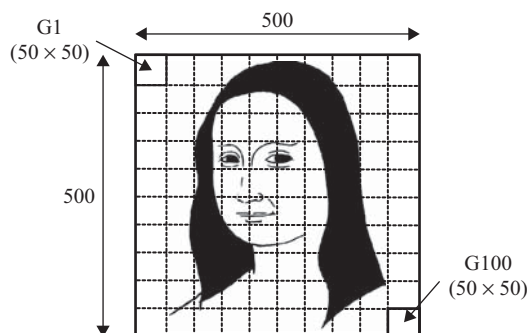


Figure 5.12. Sample 1 of unit space image.

We first find two variables, “new y_1 ” and “new y_2 ,” using the method explained in this Section, on the basis of the brightness data of the 2,500 picture elements (items) for “(a) Partial image G_1 ” shown in Table 5.25. In the same manner, for partial images G_2, \dots, G_{100} as well, we will find the two variables new y_1 and new y_2 from the 2,500 pixels, setting them up as “(b) y_1 and y_2 on the First Layer.” Note that, under circumstances free of all constraints, y_1 and y_2 would be represented as “ Y_1 ” and “ Y_2 ,” but to unmistakably distinguish them from “(c) Y_1 and Y_2 on the Second Layer,” the lower-case y was substituted here only for our immediate purposes.

Next, new Y_1 and new Y_2 are found from $y_{11}, y_{21}, \dots, y_{1100}, y_{2100}$ of “(b) y_1 and y_2 on the First Layer,” and shown as “(c) Y_1 and Y_2 on the Second Layer.” From the two variables “(c) Y_1 and Y_2 on the Second Layer,” “(d) Distance” is found, using the method discussed in this Section. What we have just introduced is an analysis method using the Multi RT Method. Note that, with this method, no problems are posed by partial populations being of varying picture element sizes or divided images having overlapped portions.

We summarize below, in Table 5.25, the flow of this analysis process with the Multi RT Method. The rectangular frames in tables (a) through (d) in the Table refer to the flow of the process in sample 1 of the Unit Space image.

Table 5.25. Flow of analysis with the Multi RT Method

(a) Partial image G_1 (left) and partial image G_{100} (right)									
No.	1	2	...	2500	No.	1	2	...	2500
1	$x_{11\ 1}$	$x_{12\ 1}$...	$x_{1\ 2500\ 1}$	1	$x_{11\ 100}$	$x_{12\ 100}$...	$x_{1\ 2500\ 100}$
2	$x_{21\ 1}$	$x_{22\ 1}$...	$x_{2\ 2500\ 1}$	2	$x_{21\ 100}$	$x_{22\ 100}$...	$x_{2\ 2500\ 100}$
...
n	$x_{n1\ 1}$	$x_{n2\ 1}$...	$x_{n\ 2500\ 1}$	n	$x_{n1\ 100}$	$x_{n2\ 100}$...	$x_{n\ 2500\ 100}$
Average	$\bar{x}_{1\ 1}$	$\bar{x}_{2\ 1}$...	$\bar{x}_{2500\ 1}$	Average	$\bar{x}_{1\ 100}$	$\bar{x}_{2\ 100}$...	$\bar{x}_{2500\ 100}$

(b) y_1 and y_2 on the First Layer

	G_1		G_2		...	G_{100}	
No.	$y_{1\ 1}$	$y_{2\ 1}$	$y_{1\ 2}$	$y_{2\ 2}$...	$y_{1\ 100}$	$y_{2\ 100}$
1	$y_{11\ 1}$	$y_{12\ 1}$	$y_{11\ 2}$	$y_{12\ 2}$...	$y_{11\ 100}$	$y_{12\ 100}$
2	$y_{21\ 1}$	$y_{22\ 1}$	$y_{21\ 2}$	$y_{22\ 2}$...	$y_{21\ 100}$	$y_{22\ 100}$
...
n	$y_{n1\ 1}$	$y_{n2\ 1}$	$y_{n1\ 2}$	$y_{n2\ 2}$...	$y_{n1\ 100}$	$y_{n2\ 100}$
Average	$\bar{y}_{1\ 1}$	$\bar{y}_{2\ 1}$	$\bar{y}_{1\ 2}$	$\bar{y}_{2\ 2}$...	$\bar{y}_{1\ 100}$	$\bar{y}_{2\ 100}$

(c) Y_1 and Y_2 on the Second

No.	Y_1	Y_2
1	Y_{11}	Y_{12}
2	Y_{21}	Y_{22}
...
n	Y_{n1}	Y_{n2}
Average	Y_1	Y_2

(d) Distance

No.	$MD(D^2)$
1	D_1^2
2	D_2^2
...	...
n	D_n^2

