

## CASE 94

# Applications of Linear and Nonlinear Regression Equations for Engineering

**Abstract:** Experiments are normally conducted following the layout of design of experiments. But in many cases, data have existed without following an experimental layout. These data are not orthogonal between factors. For such cases, experimental regression analysis can be used as a powerful tool to establish an equation that will explain the relationship between result and factors.

This study concerns linear and nonlinear cases using experimental regression analysis. The former includes an example of chemical reaction and the case using partial data from an orthogonal array, which were not used in the previous analysis to establish an equation. The latter includes the estimation of the volume of trees and the starting temperature of room heaters.

### 1. $L_{16}$ Experiment with Supplemental Characteristics

To improve the yield of a chemical product, an experiment was planned including factors  $A$  and  $B$  with four levels and factors  $C, D, E, F, G, H$  with two levels, assigned to an  $L_{16}$  orthogonal array. Beside those factors, an uncontrollable factor,  $x$ , which probably affects yield, the objective characteristic, was observed. Therefore, the magnitude of  $x$  was observed beside yield,  $y$ , for each experiment. The layout and results of this experiment are shown in Table 1.

#### Estimation of Factorial Effects Using Experimental Regression Analysis

Data with supplementary characteristics are generally analyzed using the analysis of covariance. But similar to regression analysis, it often occurs with some problems. In addition, calculation is tedious when there are many supplementary characteristics. In an experimental regression analysis, the following equation is used:

$$y = m + z_2x_1 + a_3x_2 + a_4x_3 + b_2x_4 + b_3x_5 + b_4x_6 + c_2x_7 + d_2x_8 + e_2x_8 + e_2x_9 + f_2x_{10} + g_2x_{11} + h_2x_{12} + ax \quad (1)$$

After estimating parameters, the factorial effects can be calculated. The variables are set as follows and shown in Table 2:

- $x_1$ : If level 2 is selected for  $A$ ,  $x$  is set to 1, others are set to 0.
- $x_2$ : If level 3 is selected for  $A$ ,  $x$  is set to 1, others are set to 0.
- $x_3$ : If level 4 is selected for  $A$ ,  $x$  is set to 1, others are set to 0.
- $x_4$ : If level 2 is selected for  $B$ ,  $x$  is set to 1, others are set to 0.
- $x_5$ : If level 3 is selected for  $B$ ,  $x$  is set to 1, others are set to 0.
- $x_6$ : If level 4 is selected for  $B$ ,  $x$  is set to 1, others are set to 0.

**Table 1**  
Layout and results of experiment

Row	Column											<i>x</i>	<i>y</i>
	<i>A</i> 123	<i>B</i> 48-12	<i>C</i> 5	<i>D</i> 6	<i>e</i> 7	<i>e</i> 9	<i>E</i> 10	<i>F</i> 11	<i>G</i> 13	<i>H</i> 14	<i>e</i> 15		
1	1	1	1	1	1	1	1	1	1	1	1	138	54
2	1	2	1	1	1	2	2	2	2	2	2	120	55
3	1	3	2	2	2	1	1	1	2	2	2	95	24
4	1	4	2	2	2	2	2	2	1	1	1	95	41
5	2	1	1	2	2	1	2	2	1	2	2	105	66
6	2	2	1	2	2	2	1	1	2	1	1	90	49
7	2	3	2	1	1	1	2	2	2	1	1	129	58
8	2	4	2	1	1	2	1	1	1	2	2	130	96
9	3	1	2	1	2	2	2	2	2	1	2	130	63
10	3	2	2	1	2	1	2	1	1	2	1	124	73
11	3	3	1	2	1	2	1	2	1	2	1	102	31
12	3	4	1	2	1	1	2	1	2	1	2	95	34
13	4	1	2	2	1	2	2	1	2	2	1	104	91
14	4	2	2	2	1	1	1	2	1	1	2	123	54
15	4	3	1	1	2	2	2	1	1	1	2	119	50
16	4	4	1	1	2	1	1	2	2	2	1	95	61

$x_7$ : If level 2 is selected for *C*, *x* is set to 1, others are set to 0.

$x_8$ : If level 2 is selected for *D*, *x* is set to 1, others are set to 0.

$x_9$ : If level 2 is selected for *E*, *x* is set to 1, others are set to 0.

$x_{10}$ : If level 2 is selected for *F*, *x* is set to 1, others are set to 0.

$x_{11}$ : If level 2 is selected for *G*, *x* is set to 1, others are set to 0.

$x_{12}$ : If level 2 is selected for *H*, *x* is set to 1, others are set to 0.

*x*: Supplementary characteristic.

*m*: estimate when all variables are at the first level

$a_2$ : estimate of  $\bar{A}_2 - \bar{A}_1$

$a_3$ : estimate of  $\bar{A}_3 - \bar{A}_1$

$a_4$ : estimate of  $\bar{A}_4 - \bar{A}_1$

$b_2$ : estimate of  $\bar{B}_2 - \bar{B}_1$

$b_3$ : estimate of  $\bar{B}_3 - \bar{B}_1$

$b_4$ : estimate of  $\bar{B}_4 - \bar{B}_1$

$c_2$ : estimate of  $\bar{C}_2 - \bar{C}_1$

$d_2$ : estimate of  $\bar{D}_2 - \bar{D}_1$

$e_2$ : estimate of  $\bar{E}_2 - \bar{E}_1$

$f_2$ : estimate of  $\bar{F}_2 - \bar{F}_1$

Following are definitions for *m*,  $a_2$ , ...,  $h_2$ , and *a*:

**Table 2**  
Integer variables

Row	Column												x	y
	x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>	x <sub>4</sub>	x <sub>5</sub>	x <sub>6</sub>	x <sub>7</sub>	x <sub>8</sub>	x <sub>9</sub>	x <sub>10</sub>	x <sub>11</sub>	x <sub>12</sub>		
1	0	0	0	0	0	0	0	0	0	0	0	0	138	54
2	0	0	0	1	0	0	0	0	1	1	1	1	120	55
3	0	0	0	0	0	0	1	1	0	0	1	1	95	24
4	0	0	0	0	0	1	1	1	1	1	0	0	95	41
5	1	0	0	0	0	0	0	1	1	1	0	1	105	66
6	1	0	0	1	0	0	0	1	0	0	1	0	90	49
7	1	0	0	0	0	0	1	0	1	1	1	0	129	58
8	1	0	0	0	0	1	1	0	0	0	0	1	130	96
9	0	1	0	0	0	0	1	0	0	1	1	0	130	63
10	0	1	0	1	0	0	1	0	1	0	0	1	124	73
11	0	1	0	0	0	0	0	1	0	1	0	1	102	31
12	0	1	0	0	0	1	0	1	1	0	1	0	95	34
13	0	0	1	0	0	0	1	1	1	0	1	1	104	91
14	0	0	1	1	0	0	1	1	0	1	0	0	123	54
15	0	0	1	0	0	0	0	0	1	0	0	0	119	50
16	0	0	1	0	0	1	0	0	0	1	1	1	95	61

$g_2$ : estimate of  $\bar{G}_2 - \bar{G}_1$

$h_2$ : estimate of  $\bar{H}_2 - \bar{H}_1$

$a$ : change of yield,  $y$ , by unit change of  $x$

Table 3 shows the initial and converged values at the fifth iteration. The initial level intervals were set wide. These can be narrower if engineering knowledge is utilized. Using the second levels of the fifth iteration yields

$$m = \bar{y} - (a_2x_1 + a_3x_2 + \dots + h_2x_{12} + a\bar{x}) = -26 \quad (2)$$

Comparison was made as shown in Table 4. For verification it was desirable to use the data that were not used for the estimation of parameters. But in this case, all 16 data points are used for verification. The average of differences squared was calculated as 1412.60. This corresponds to error variance.

## 2. Partial Data from an Experiment Using an Orthogonal Array

In an experiment using an  $L_{18}$  array, two-level factor  $A$  was assigned to the first column and three-level factors  $B$ ,  $C$ ,  $D$ ,  $E$ , and  $F$  to columns 2 to 6, respectively. The smaller-the-better characteristic was used, calculated from the noise factors assigned to the outer array. Factors  $A$ ,  $D$ , and  $F$  are significant. Using these factors to estimate the SN ratio under the optimum configuration, the SN ratio was  $-36.97$  dB, and that under current conditions was  $-38.61$ . There is an improvement of about 1.6 dB.

The estimate under current conditions was close to the result of number 1, where all factors are at their first level (Table 5). The SN ratios of number 6 or 8 are better than the one under the optimum, suggesting that the SN ratio under the optimum might be better. Although the SN ratio analysis of

**Table 3**  
Initial values and values after convergence

Parameter	Initial Level			Level after Fifth Iteration		
	1	2	3	1	2	3
$a_2$	0.0	16.0	32.0	16.0	24.0	32.0
$a_3$	0.0	16.0	32.0	0.0	8.0	16.0
$a_4$	0.0	16.0	32.0	16.0	24.0	32.0
$b_2$	-32.0	-16.0	0.0	-16.0	-8.0	0.0
$b_3$	-32.0	-16.0	0.0	-32.0	-16.0	0.0
$b_4$	-8.0	0.0	8.0	-8.0	0.0	8.0
$c_2$	8.0	0.0	8.0	-8.0	0.0	8.0
$d_2$	-8.0	0.0	8.0	-8.0	0.0	8.0
$e_2$	-8.0	0.0	8.0	-8.0	0.0	8.0
$f_2$	-8.0	0.0	8.0	-8.0	0.0	8.0
$g_2$	-8.0	0.0	8.0	-8.0	0.0	8.0
$h_2$	-8.0	0.0	8.0	4.0	6.0	8.0
$a$	0.0	0.8	1.6	0.4	0.6	0.8

$L_{18}$  is not shown here, the confidence interval was quite wide from the analysis of variance. In such a case it is natural that the researcher wants to have a better estimate.

It was noted from observations that there is a big difference between the SN ratios from numbers 1 to 9 and those from numbers 10 to 18. In addition, the SN ratios under  $A_2$  are close to each other. In other words, the effects of  $B$ ,  $C$ ,  $D$ ,  $E$ ,  $F$ , and  $G$  under  $A_2$  are insignificant. But it is not clear whether there are problems in the level setting for those factors or whether they are truly insignificant. Since the effects of factors  $B$  to  $G$  are different at different levels of  $A$ , which indicates the existence of interactions, the error calculated may become unsatisfactorily large.

Since the results under  $A_1$  were better, it may be that the effects of  $B$  to  $G$  can be estimated from the results under  $A_1$ . As seen in the upper half of  $L_{18}$ , the array of columns 2, 3, 6, and 7 are identical to the  $L_9$  array. Even for the case when there are other factors assigned to other columns in the upper half of  $L_{18}$ , if their effects were small, analysis could be

made from the upper half, although it is not a general practice.

If we want to estimate the effects from the upper half of  $L_{18}$ , it is necessary that the factors be orthogonal to each other. To analyze the data that are not orthogonal, experimental regression analysis is useful.

#### Preparation of Data for Regression Analysis

Table 5 shows the SN ratios of experiments 1 to 9. To analyze the effects of  $B$  through  $G$ , those factors were replaced with integer-type variables and the following equation was used:

$$y = m + b_2x_1 + b_3x_2 + c_2x_3 + c_3x_4 + d_2x_5 + d_3x_6 + e_2x_7 + e_3x_8 + f_2x_9 + f_3x_{10} \quad (3)$$

Variables were set as follows (Table 6):

$x_1$ : If level 2 is selected for  $B$ ,  $x_1$  is set to 1, others are set to 0.

**Table 4**

Comparison between estimation and observation

No.	Observation	Estimation	(Est.) – (Obs.)
1	54.0	52.6	-1.4
2	55.0	41.4	-13.6
3	24.0	39.4	15.4
4	41.0	34.4	-5.6
5	66.0	67.4	1.4
6	49.0	33.4	-15.6
7	58.0	73.0	15.0
8	96.0	93.4	-2.6
9	63.0	49.4	-13.6
10	73.0	75.0	2.0
11	31.0	42.2	11.2
12	34.0	34.4	1.4
13	91.0	75.0	-16.0
14	54.0	54.6	0.6
15	50.0	77.0	27.0
16	61.0	55.4	-5.6

$x_2$ : If level 3 is selected for *B*,  $x$  is set to 1, others are set to 0.

$x_3$ : If level 2 is selected for *C*,  $x$  is set to 1, others are set to 0.

$x_4$ : If level 3 is selected for *C*,  $x$  is set to 1, others are set to 0.

$x_5$ : If level 2 is selected for *D*,  $x$  is set to 1, others are set to 0.

$x_6$ : If level 3 is selected for *D*,  $x$  is set to 1, others are set to 0.

$x_7$ : If level 2 is selected for *E*,  $x$  is set to 1, others are set to 0.

$x_8$ : If level 3 is selected for *E*,  $x$  is set to 1, others are set to 0.

$x_9$ : If level 2 is selected for *F*,  $x$  is set to 1, others are set to 0.

$x_{10}$ : If level 3 is selected for *F*,  $x$  is set to 1, others are set to 0.

The coefficients show the following contents:

**Table 5**

SN ratios

Row	Column								dB
	A 1	B 2	C 3	D 4	E 5	F 6	e 7	e 8	
1	1	1	1	1	1	1	1	1	-39.79
2	1	1	2	2	2	2	2	2	-42.74
3	1	1	3	3	3	3	3	3	-46.52
4	1	2	1	1	2	2	3	3	-44.33
5	1	2	2	2	3	3	1	1	-46.65
6	1	2	3	3	1	1	2	2	-34.92
7	1	3	1	2	1	3	2	3	-43.75
8	1	3	2	3	2	1	3	1	-32.81
9	1	3	3	1	3	2	1	2	-40.98

**Table 6**  
Placement of factors in Table 5 as integer-type variables

<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$	<b>y</b>
<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	$B_2$	$B_3$	$C_2$	$C_3$	$D_2$	$D_3$	$E_2$	$E_3$	$F_2$	$F_3$	
1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	-39.79
1	2	2	2	2	0	0	1	0	1	0	1	0	1	0	-42.74
1	3	3	3	3	0	0	0	1	0	1	0	1	0	1	-46.52
2	1	1	2	2	1	0	0	0	0	0	1	0	1	0	-44.33
2	2	2	3	3	1	0	1	0	1	0	0	1	0	1	-46.55
2	3	3	1	1	1	0	0	1	0	1	0	0	0	0	-34.92
3	1	2	1	3	0	1	0	0	1	0	0	0	0	1	-43.75
3	2	3	2	1	0	1	1	0	0	1	1	0	0	0	-32.81
3	3	1	3	2	0	1	0	1	0	0	0	1	1	0	-40.98

$m$ : estimate of  $(A_1)B_1C_1D_1E_1F_1G_1$

$b_2$ : estimate of  $\bar{B}_2 - \bar{B}_1$

$b_3$ : estimate of  $\bar{B}_3 - \bar{B}_1$

$c_2$ : estimate of  $\bar{C}_2 - \bar{C}_1$

$c_3$ : estimate of  $\bar{C}_3 - \bar{C}_1$

$d_2$ : estimate of  $\bar{D}_2 - \bar{D}_1$

$d_3$ : estimate of  $\bar{D}_3 - \bar{D}_1$

$e_2$ : estimate of  $\bar{E}_2 - \bar{E}_1$

$e_3$ : estimate of  $\bar{E}_3 - \bar{E}_1$

$f_2$ : estimate of  $\bar{F}_2 - \bar{F}_1$

$f_3$ : estimate of  $\bar{F}_3 - \bar{F}_1$

$$y = -38.71 + x_1 + 2.5x_2 + x_3 + 1.5x_4 + x_5 + x_6 + x_7 - x_8 - 7x_9 - 9x_{10} \quad (5)$$

Table 8 shows the comparison by putting the values in Table 7. The error variance was calculated as 0.64. Figure 1 shows the response graphs where the effect is the gain from level 1.

### Estimation of the Optimal Configuration

Since the optimum configuration is  $B_3C_3D_2(D_3)E_2F_1$ ,

$$y = -38.71 = x_1 + 2.5x_2 + x_3 + 1.5x_4 + x_5 + x_6 + x_7 - x_8 - 7x_9 - 9x_{10} \quad (6)$$

where  $x_1 = 0, x_2 = 1, x_3 = 0, x_4 = 1, x_5 = 1, x_6 = 0, x_7 = 1, x_8 = 0, x_9 = 0,$  and  $x_{10} = 0$ .

### 3. Other Applications

#### Ranges of Initial Values

The initial values of  $b_2, \dots, f_3$  are determined based on engineering knowledge (Table 7). Setting the results of level 2 of the eighth iteration,

$$m = y - (b_2x_1 + b_3x_2 \dots + f_3x_{10}) = -38.71 \quad (4)$$

$y$  is calculated as

In the section above, only the optimum configuration was estimated. But it is possible to estimate any combination. Sometimes there are incomplete (missing) data in an orthogonal experiment, and sequential analysis is conducted followed by analysis of variance. Even in such a case, the estimation of missing combination(s) can be made by calculating

**Table 7**  
Initial and converged values of parameters

Coefficient	Level					
	Initial Three levels			Conveyed Eighth Iteration		
	1	2	3	1	2	3
$b_2: \bar{B}_2 - \bar{B}_1$	0	2	4	0	1	2
$b_3: \bar{B}_3 - \bar{B}_1$	0	2	4	2	2.5	3
$c_2: \bar{C}_2 - \bar{C}_1$	-2	0	2	0	1	2
$c_3: \bar{C}_3 - \bar{C}_1$	-2	0	2	1	1.5	2
$d_2: \bar{D}_2 - \bar{D}_1$	-2	0	2	0	1	2
$d_3: \bar{D}_3 - \bar{D}_1$	-2	0	2	0	1	2
$e_2: \bar{E}_2 - \bar{E}_1$	-2	0	2	0	1	2
$e_3: \bar{E}_3 - \bar{E}_1$	-2	0	2	-2	-1	0
$f_2: \bar{F}_2 - \bar{F}_1$	-8	-4	0	-8	-7	-6
$f_3: \bar{F}_3 - \bar{F}_1$	-16	-8	0	-10	-9	-8

the coefficients using experimental regression analysis from the rest of the data.

In the experiments of manufacturing areas, data are collected based on the layout of experimental design. But in some cases, we want to utilize the existing data from observation prior to experimentation. In most cases, there is no orthogonality

**Table 8**  
Comparison between estimation and observation

No.	Observation	Estimation	(Est.) - (Obs.)
1	-39.79	-38.71	1.08
2	-42.74	-42.71	0.03
3	-46.52	-46.21	0.31
4	-44.33	-43.71	0.62
5	-46.55	-45.71	0.84
6	-34.92	-35.21	-0.29
7	-43.75	-44.21	-0.46
8	-32.81	-33.21	-0.40
9	-40.98	-42.71	-1.73

between factors. For such cases, experimental regression analysis can be used as a powerful tool.

#### 4. Estimation of the Volume of Trees

The volume of a tree is estimated from chest-high diameter,  $D$ , and height,  $H$ . The following equation is generally used:

$$y = a_1 D^{a_2} H^{a_3} \quad (7)$$

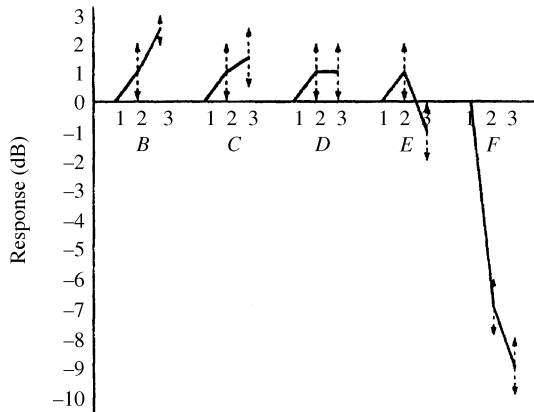
If the coefficients  $a_1$ ,  $a_2$ , and  $a_3$  are known, the volume of a tree is estimated from chest height and tree height.

To determine  $a_1$ ,  $a_2$ , and  $a_3$ , 52 trees were cut down and the values,  $D$  and  $H$ , and volume,  $y$ , were measured (Table 9).

Although it is possible to conduct the logarithmic transformation, then estimate the coefficients by linear regression analysis, nonlinear equation (7) is used without transformation in this case.

##### Sequential Approximation

Since  $D$  and  $H$  are diameter and height, respectively, it looks like  $D$  is close to the power of 2, and  $H$  is around 1. In other words,  $a_2$  is around 2 and  $a_3$  is



**Figure 1**  
Response graphs of partial data

around 1. Considering the units of  $D$  being in centimeters,  $H$  being in meters, and  $y$  being in square meters, the unit of  $a_1$  is probably around 0.0001. The initial values are set as shown in Table 10.

**Converged Values and Their Evaluation**

After conducting sequential approximation, it converged after nine iterations to obtain the results shown in Table 11. Volume,  $y$ , is given by the following equation using the values of the second level:

$$y = 0.0001096D^{1.839}H^{0.8398} \tag{8}$$

The total of residuals squared,  $S_e$ , is 0.02423, and the standard deviation is 0.021.

Next, the average of residual error variation was calculated from the data in Table 12, the observa-

**Table 9**  
Observational results for trees

No.	$D$ (cm)	$H$ (m)	$y$ (m <sup>3</sup> )
1	36.3	22.50	1.161
2	24.5	18.83	0.461
3	26.3	17.35	0.535
4	29.5	19.22	0.701
⋮	⋮	⋮	⋮
52	14.5	14.37	0.124

**Table 10**  
Initial coefficients

	1	2	3
$a_1$	0.00005	0.00010	0.00015
$a_2$	1.5	2.0	2.5
$a_3$	0.5	1.0	1.5

tions of 10 trees that were not used in the estimation of parameters. It was calculated from the average of the differences between estimation and observation squared.

$$V = \frac{1}{10[(-0.003)^2 + 0.060^2 + \dots + 0.009^2]} = 0.00554 \tag{9}$$

The standard deviation was 0.023, showing good reproducibility of conclusions for the data that were not used for the equation of estimation, shown in Tables 11, 12, and 13.

**Comparison with Logarithmic Transformation**

As described above, equation (7) can be converted into a linear equation by logarithmic transformation:

$$\log y = \log a_1 + a_2 \log D + a_3 \log H \tag{10}$$

If  $a_2$  and  $a_3$  are known,  $\log a_1$  is calculated as a constant. The following equation, which was obtained from the least squares method, does not have any technical contradictions and is considered as one of the solutions.

**Table 11**  
Converged coefficients after ninth iteration

	1	2	3
$a_1$	0.0001094	0.0001096	0.0001098
$a_2$	1.8379	1.8398	1.8418
$a_3$	0.8387	0.8398	0.8418



**Table 12**

Data for verification and estimated values

No.	<i>D</i> (cm)	<i>H</i> (m)	<i>y</i> (m <sup>3</sup> )	$\hat{y}$ (m <sup>3</sup> )	$\hat{y} - y$
1	46.3	21.90	1.701	1.698	-0.003
2	25.8	20.65	0.491	0.551	0.060
3	27.7	17.60	0.566	0.549	-0.017
4	23.6	20.05	0.473	0.456	-0.017
5	21.2	14.50	0.272	0.285	0.013
6	20.2	17.90	0.337	0.312	-0.025
7	18.1	17.33	0.264	0.248	-0.016
8	14.8	13.51	0.128	0.139	0.011
9	14.4	13.83	0.126	0.135	0.011
10	11.8	12.11	0.075	0.083	0.008

$$\log y = -4.08817 + 1.87475 \log D + 0.902218 \log H \quad (11)$$

A verification was made similarly to Table 12 using this equation, and the average squared error was calculated as

$$V = \frac{1}{10(0.052^2 + 0.064^2 + \dots + 0.004^2)} = 0.000864 \quad (12)$$

This is slightly larger than the result from equation (9).

**Table 13**

Data for verification and estimated values

No.	<i>D</i> (cm)	<i>H</i> (m)	<i>y</i> (m <sup>3</sup> )	$\hat{y}$ (m <sup>3</sup> )	$\hat{y} - y$
1	46.3	21.90	1.701	1.753	0.052
2	24.8	20.65	0.491	0.555	0.064
3	27.7	17.60	0.566	0.549	-0.017
4	23.6	20.05	0.473	0.458	-0.015
5	21.2	14.50	0.272	0.279	0.007
6	20.2	17.90	0.337	0.309	-0.028
7	18.1	17.33	0.264	0.244	-0.020
8	14.8	13.51	0.128	0.134	0.006
9	14.4	13.83	0.126	0.130	0.006
10	11.8	12.11	0.075	0.079	0.004

Comparing the far-right columns in Tables 12 and 13, it can be seen that there is a big difference between the two tables on tree 1. The volume of this tree is larger than the volume of others. Generally, the absolute error becomes larger after logarithmic transformation. In the case of estimating the volume of the trees in forests, a 3% error of larger trees with a volume of such as 1.5 m<sup>3</sup> is more important than a 10% error of smaller trees with a volume of 0.2 m<sup>3</sup>. Therefore, using logarithmic transformation to estimate  $a_2$  or  $a_3$ , the weight of larger trees must be taken into consideration.

When using equation (7), the absolute error is not affected by volume, even when weight is not considered. It is therefore recommended that logarithmic transformation not be used when the loss caused by estimation error is affected by the absolute error.

## 5. Starting Temperature of Room Heaters

Most office workers begin their day at 9 a.m., and it is therefore desirable that the temperature at 9 a.m. be at the targeted room temperature in winter. To make this possible, the heater (or air conditioner in summer) must start 30 or 60 minutes earlier. Although the same period of time is used for preheating, overheating or underheating occur due to the outside temperature, sunshine, the length of time operated on the previous day, and so on, so the temperature often does not hit the target at the time that one wishes to start using the room.

Let the temperature to start heating be  $\theta_n$  and the temperature difference between outside temperature and design temperature (0°C for Tokyo and -1°C for Yokohama) to start heating be  $\Delta\theta_A$ . The room temperature,  $T$ , hours later is given from heat transfer theory by

$$\theta_r = \theta_n + (1 - e^{-\alpha t})k(1 + \beta \Delta\theta_A) \quad (13)$$

where  $\alpha$  is the coefficient of transfer,  $k$  the constant (usually set to 4.5 based on past experiments), and  $\beta$  the calibration coefficient for outer temperature (usually set to 0.031). The preheating time,  $T_0$ , with a target room temperature of  $\theta_0$  is given from equation (13) as

$$T_0 = \frac{1}{\alpha \ln\{1 - [\theta_0 - \theta_n/k(1 + \beta\Delta\theta_A)]\}} \quad (14)$$

The values of  $\alpha$ ,  $\beta$ , and  $k$  are different for different building structures and other conditions. It is important to determine those values for the development of energy-saving equipment.

### Results of Observation and Estimation for Coefficients of Function

An air-conditioning control equipment manufacturer collected the data from some buildings to estimate the coefficients mentioned in the preceding section. During the 26-day period, observation was made on the day of the week, climate, temperature difference between outer and design target ( $\Delta\theta_A$ ), room temperature when the heater started ( $\theta_n$ ), the time from starting ( $T$ ), and current room temperature ( $\theta_r$ ). Measurements were done three times a day. Some of the results are shown in Table 14.

To estimate  $\alpha$ ,  $\beta$  and  $k$  in equation (13), some of the data in Table 14 were converted into integer-type variables, to obtain 11 variables in total. Regarding the week of the day, there are holidays beside Sunday; therefore, it is better to set the variable as follows (Table 15):

- $x_1$ : Second day from holiday is set to 1, others are set to 0.
- $x_2$ : Third day from holiday is set to 1, others are set to 0.
- $x_3$ : Fourth day from holiday is set to 1, others are set to 0.
- $x_4$ : Fifth day from holiday is set to 1, other are set to 0.
- $x_5$ : Sixth day from holiday is set to 1, others are set to 0.

For example,  $x_1$  to  $x_5$  are equal to 0 for the day right after a holiday. For climate,

- $x_6$ : Cloudy is set to 1, others are set to 0.
- $x_7$ : Snow is set to 1, others are set to 0.

Therefore,  $x_6$  and  $x_7$  are equal to zero for a fine day. The variables of other items are

**Table 14**

Results of observation

No.	Day <sup>a</sup>	Climate	$\Delta\theta_A$	$\theta_r$	T (hours)	$\theta_n$
1	Tuesday <sup>b</sup>	Cloudy	4.0	20.0	1.00	18.5
2	—	—	—	21.7	2.00	—
3	—	—	—	22.5	3.00	—
4	Wednesday	Snow	0.0	21.0	1.00	19.5
5	—	—	—	22.5	2.00	—
6	—	—	—	23.0	3.00	—
7	Thursday	Cloudy	0.0	20.5	0.17	20.0
8	—	—	—	22.0	0.33	—
9	—	—	—	23.0	0.50	—
⋮	⋮	⋮	⋮	⋮	⋮	⋮
76	Monday <sup>b</sup>	Clear	2.5	22.0	0.25	21.5
77	—	—	—	22.8	0.50	—
78	—	—	—	23.3	0.75	—

<sup>a</sup>—, observation same as above since made on the same day.<sup>b</sup>Day next to a holiday when the heater was not operated.**Table 15**

Variables, including integer types for room temperature

No.	1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	0	0	1	0	4.0	20.0	1.00	18.5
2	0	0	0	0	0	1	0	4.0	21.7	2.00	18.5
3	0	0	0	0	0	1	0	4.0	22.5	3.00	18.5
4	1	0	0	0	0	0	1	0.0	21.0	1.00	19.5
5	1	0	0	0	0	0	1	0.0	22.5	2.00	19.5
6	1	0	0	0	0	0	1	0.0	23.0	3.00	19.5
⋮							⋮				
76	0	0	0	0	1	0	0	2.5	22.0	0.25	21.5
77	0	0	0	0	1	0	0	2.5	22.8	0.50	21.5
78	0	0	0	0	1	0	0	2.5	23.3	0.75	21.5

**Table 16**  
Initial and convergent values of parameters for room temperature

	Level					
	Initial Values			After 16th Iteration		
	1	2	3	1	2	3
$a_1$	0	0.8	1.6	0.3	0.4	0.5
$a_2$	0	0.8	1.6	0.7	0.75	0.8
$a_3$	0	0.8	1.6	0.7	0.75	0.8
$a_4$	0	0.8	1.6	0.4	0.5	0.6
$a_5$	0	0.8	1.6	0.4	0.45	0.5
$a_6$	-0.8	-0.4	0	-0.275	-0.2625	-0.25
$a_7$	-0.8	-0.4	0	-0.4	-0.35	-0.3
$a_8$	0.4	0.8	1.4	0.7	0.725	0.75
$a_9$	4	8	12	5.0	5.25	5.5
$a_{10}$	0	0.016	0.032	0.0	0.001	0.002

$x_8$ : $\Delta\theta_A$	$\alpha$ (2 days after a holiday) = $a_8(1 + a_1)$
$x_9$ : $\theta_r (= y)$	$\alpha$ (3 days after a holiday) = $a_8(1 + a_2)$
$x_{10}$ : $T$	⋮
$x_{11}$ : $\theta_n$	$\alpha$ (6 days after a holiday) = $a_8(1 + a_5)$

$\alpha$ , the heat transfer coefficient, seemingly has a small value if the heater was not operated the preceding day.  $\beta$  is the coefficient to calibrate the outer temperature: a constant for a particular building.  $k$  is a proportional constant for temperature rise during time passage, it is affected by climate.

The equations for  $\alpha$  and  $k$  are set as

$$\alpha = a_8(1 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5) \tag{15}$$

$$k = a_9(1 + a_6x_6 + a_7x_7) \tag{16}$$

Variable  $a_{10}$  represents  $\beta$ . Determinating  $\alpha$ ,  $\beta$ , and  $k$  therefore involves determining 10 variables,  $a_1$  to  $a_{10}$ .

Letting  $a_8$  be the value of  $\alpha$  the day after a holiday:

Letting  $a_9$  represent  $k$  for a fine day, then

$$k(\text{cloudy}) = a_9(1 + a_6)$$

$$k(\text{snow}) = a_9(1 + a_7)$$

Parameter Estimation by Experimental Regression Analysis

Letting  $\theta_r(x_9)$  be the objective variable,  $y$ :

$$y = x_{11} + (1 - e^{\alpha x_{10}})k(1 + \alpha_{10}x_8) \tag{17}$$

where

$$\alpha = a_8(1 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5) \tag{18}$$

$$k = a_9(1 + a_6x_6 + a_7x_7) \tag{19}$$

**Table 17**

Deviations and total residuals squared after the sixth iteration

No.	Deviation	Total Residuals Squared	Combination													
1	0.296194191	33.0726635	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	0.881743940E-01	28.4722355	2	2	2	2	2	2	2	2	2	2	2	2	2	1
3	-0.132098118	31.0850840	3	3	3	3	3	3	3	3	3	3	3	3	3	1
4	0.372788513E-01	29.2814816	1	1	1	1	2	2	2	2	3	3	3	3	3	1
⋮		⋮														
35	-0.146478391E-01	27.7087209	2	1	2	3	1	3	1	2	3	3	1	2	3	
36	0.149078921	29.7143803	3	2	3	1	2	1	2	3	1	1	2	3	3	

The initial parameter values and the ones after convergence are shown in Table 16.

For sequential approximation, the data of twenty-three out of twenty-six days were used. Some of the deviations and the total residuals squared are shown in Table 17. Using the second level, the values of  $a$ ,  $b$ , and  $k$  are shown in Table 18.

The proportional coefficient of heat transfer ( $\alpha$ ) for the day after a holiday is about 30% smaller than on other days. The result that coefficients three and four days after a holiday are larger than those five

and six days after a holiday does not make sense. Probably, only a comparison between the day next to a holiday and other days has actual meaning. The calibration coefficient for outer temperature,  $\beta$ , is 0.001, showing that the effect of outer temperature is small.

The constant for temperature rise due to time passage,  $k$ , is affected by climate. The value is about 70% of that on a fine day. On a bad weather day, the temperature becomes lower, with other conditions unchanged.

**Table 18**

Values of  $\alpha$ ,  $\beta$ , and  $k$

Parameter	$x$	Relation to Unknown Parameter $a_1$	Value
$\alpha$	Day after holiday	$a_8$	0.725
	2 days after holiday $x_1$	$a_8(1 + a_1)$	$0.725(1 + 0.4) = 1.015$
	3 days after holiday $x_2$	$a_8(1 + a_2)$	$0.725(1 + 0.75) = 1.26875$
	4 days after holiday $x_3$	$a_8(1 + a_3)$	$0.725(1 + 0.75) = 1.26875$
	5 days after holiday $x_4$	$a_8(1 + a_4)$	$0.725(1 + 0.5) = 1.0875$
	6 days after holiday $x_5$	$a_8(1 + a_5)$	$0.725(1 + 0.45) = 1.05125$
$\beta$		$a_{10}$	0.001
$k$	Clear	$a_9$	5.25
	Cloudy	$a_9(1 + a_6)$	$5.25(1 - 0.2625) = 3.871875$
	Snow	$a_9(1 + a_7)$	$5.25(1 - 0.35) = 3.4125$

### Estimation of Preheating Time

Preheating time ( $T_0$ ) can be estimated from the target temperature ( $\theta_0$ ) by putting the parameters above into equation (14).

There might be a tendency to use equation (14) as the regression equation since what we need is the time to preheat,  $T_0$ . However, what was measured in this study was room temperature against the time

after heating started. In other words, time passage is the cause and room temperature is the effect. What can be used for adjustment is time, and the result is room temperature. It must be noted as a general rule that a regression equation is supposed to express a result,  $y$ , by the cause,  $x$ .

---

*This case study is contributed by Genichi Taguchi.*