32 Two-Way Layout

32.1. Introduction 552
32.2. Factors and Levels 552
32.3. Analysis of Variance 555

32.1. Introduction

Let us discuss how to determine a way of increasing the production yield of a chemical product. Many factors might affect yield of the product; in this case we only discuss an experiment designed for determining the catalyst quantity and synthesis temperature to obtain a high yield. This chapter is based on Genichi Taguchi et al., *Design of Experiments*. Tokyo: Japanese Standards Association, 1973.

32.2. Factors and Levels

Causes of a given result in an experiment are called *factors*. Factors such as temperature and catalyst quantity at the stage of synthesis are denoted by A and B, respectively. If the relationships between factor A, factor B, and yield are determined, we can decide the temperature and catalyst quantity that will result in a higher yield.

From our past experience, we can roughly estimate a temperature range that will give the highest yield: for example, a temperature range of 200 to 300°C. In addition, we may know that the quantity of a catalyst that will give us the highest yield is in the range 0.2 to 0.8%. Although such ranges are known roughly in many cases, the yield may vary substantially within the ranges above; for this reason we need an experiment that will determine precisely the best temperature and catalyst quantity.

To determine the relationships among yield, temperature, or catalyst quantity, the conditions of temperature and catalyst quantity are varied, and the resulting relationships are plotted in graphs. This particular type of experiment is usually carried out as follows: First, the ranges of temperature and catalyst quantity are fixed and two or three temperatures and catalyst quantities within the ranges are selected. For example, when the temperature range is between 200 and 300°C, the following temperatures may be selected: A_1 = 200°C, A_2 = 225°C, A_3 = 250°C, A_4 = 275°C, and A_5 = 300°C.

After the temperature range is fixed, the remaining problem is to decide the number of points within the range to be used. The number of points needed depends on the complexity of the curve of temperature, which is plotted against the yield. Like the experiment described above, three to five points are usually chosen because the curve of temperature and the yield normally has a mountainous shape or a simple smooth curve. If there is a possibility for a curve with two or more mountains, more temperature points must be chosen.

Suppose that we have the range 200 to 300°C and take five points. In this case we define the temperature at five different levels at equal intervals. In the same way, levels are chosen for the catalyst quantity factor. If it is known before the experiment begins that the catalyst quantity must be increased when the temperature is increased, the range of catalyst quantity should be adjusted for each temperature level.

This is not likely to happen in the relationship between temperature and catalyst quantity, but obviously does happen in the case of temperature and reaction time. This particular phenomenon often happens with various related factors of chemical reactions: for example, the reaction of hydrogen and oxygen to produce water,

$$2H_2 + O_2 = 2H_2O$$

where two molecules of hydrogen and one molecule of oxygen react. Consequently, the quantity of oxygen is in the neighborhood of half the chemical equivalent of hydrogen.

To find the effect caused by different flows per unit of time, suppose that the quantity of H_2 is varied as A_1, A_2, and A_3. The quantity of oxygen is then set at the following three levels:

B_1: theoretical amount (chemical equivalent)

B_2: 1.02 times B_1 (2% more than the theoretical amount)

B_3: 1.04 times B_1 (4% more than the theoretical amount)

Why are levels B_2 and B_3 necessary? To minimize the productivity loss caused by the amount of unreacted material, we want to investigate how much excess oxygen can be added to increase profitability maximally, since oxygen happens to be cheaper than hydrogen. (The purpose of setting three levels for the quantity of hydrogen is to investigate the effect of hydrogen flow.)

In this particular experiment, we assume that it is not necessary to forecast the temperature range for each distinct catalyst quantity; therefore, the range of catalyst quantity can be independent of temperature. Suppose that the range is 0.2 to 0.8% and that four levels are chosen at equal intervals. Thus, B_1 = 0.2%,

$B_2 = 0.4\%$, $B_3 = 0.6\%$, and $B_4 = 0.8\%$. There are now 20 combinations of factors A and B as shown in Table 32.1.

The order of experimenting with each combination is determined by drawing lots. Random order is needed because other factors may also be influencing the yield. Uncontrollable variations of raw or subsidiary materials or the catalyst itself may influence the results. Errors in measurement of the yield may also create biased results. There will be differences in the experiments, even though the conditions are controlled as much as possible.

Suppose that the experiments were conducted and the yields obtained are those shown in Table 32.1. The following questions must now be answered:

Table 32.1
Experiment with a two-way layout

Experiment No.	Level of *A*	Level of *B*	Combination of *A* and *B*	Actual Condition of Experiment: Temperature (°C)	Actual Condition of Experiment: Catalyst Quantity (%)	Data in Yield (%)
1	1	1	A_1B_1	200	0.2	64
2	1	2	A_1B_2	200	0.4	65
3	1	3	A_1B_3	200	0.6	76
4	1	4	A_1B_4	200	0.8	64
5	2	1	A_2B_1	225	0.2	67
6	2	2	A_2B_2	225	0.4	81
7	2	3	A_2B_3	225	0.6	82
8	2	4	A_2B_4	225	0.8	91
9	3	1	A_3B_1	250	0.2	76
10	3	2	A_3B_2	250	0.4	81
11	3	3	A_3B_3	250	0.6	88
12	3	4	A_3B_4	250	0.8	90
13	4	1	A_4B_1	275	0.2	76
14	4	2	A_4B_2	275	0.4	84
15	4	3	A_4B_3	275	0.6	83
16	4	4	A_4B_4	275	0.8	92
17	5	1	A_5B_1	300	0.2	73
18	5	2	A_5B_2	300	0.4	80
19	5	3	A_5B_3	300	0.6	84
20	5	4	A_5B_4	300	0.8	91

1. How much influence against the yield will be caused when the temperature or catalyst quantity changes?
2. What is the appearance of the curve showing the relation of temperature and yield or catalyst quantity and yield?
3. Which temperature and what catalyst quantity will give the highest yield?

An investigation of these questions follows.

32.3. Analysis of Variance

First, the decomposition of variation has to be derived. We want to know how the yield changes when the temperature or catalyst quantity changes within the experimental range. In other words, we want to show the composite effects of temperature and catalyst quantity on yield and in such a way answer the three questions posed earlier.

If every experiment in Table 32.1 resulted in 64% yield, this would indicate that neither temperature nor the quantity of the catalyst affects yield within that particular factor range, or it would be shown that neither materials nor the measuring skill caused the variance.

However, the average yield is 64% instead of 100%; this is a big problem. This was caused by the factors that were kept fixed during the period of the experiment. That is, the average value of experimental data reflects the influence of nonvariable factors. The average might not have been 64% had the type of catalyst or quantity of materials in the reaction been changed.

The problem of discussing the influence of those fixed factors belongs to the technologists who work in this specialized field; it is normally excluded from the analysis of data. For this reason, the correction factor is

$$\mathrm{CF} = \frac{(\text{total of all data})^2}{20} \tag{32.1}$$

which is subtracted from the total sum of the squares to get the total variation, S_T; then S_T is decomposed. First, a working mean, 80, is subtracted from each result in Table 32.1. The data subtracted are shown in Table 32.2.

Table 32.2
Subtracted data

	B_1	B_2	B_3	B_4	Total
A_1	−16	−15	−4	−16	−51
A_2	−13	1	2	11	1
A_3	−4	1	8	10	15
A_4	−4	4	3	12	15
A_5	−7	0	4	11	8
Total	−44	−9	13	28	−12

To verify the accuracy of the table, the following totals are calculated: first, a total is made for each row and each column; next, the totals for both the rows and columns are brought to grand totals; these grand totals must then be equal. The correction factor is

$$\mathrm{CF} = \frac{(-12)^2}{20} = 7 \qquad (f = 1) \tag{32.2}$$

The total amount of variation S_T is

$$\begin{aligned} S_T &= (-16)^2 + (-15)^2 + \cdots + 11^2 - \mathrm{CF} \\ &= 1600 - 7 = 1593 \qquad (f = 19) \end{aligned} \tag{32.3}$$

Next, calculate the effect of A, which is shown by S_A.

$$S_A = \tfrac{1}{4}[(-51)^2 + 1^2 + 15^2 + 15^2 + 8^2] - 7 = 772 \qquad (f = 4) \tag{32.4}$$

The following is the reason that S_A can be calculated by equation (32.4).

In Table 32.2, the effect of temperature, which is factor A, is included in each condition for each level of catalyst quantity: B_1, B_2, B_3, and B_4. Accordingly, the effect of A must be obtained from each level of B and the effects added together. This is especially important when the effects of temperature are not the same under different levels of catalyst quantity.

When the effects of temperature A are not the same under varying amounts of catalyst, it is known that interaction exists between temperature A and catalyst quantity B. This interaction is written as $A \times B$. Strictly speaking, interaction exists in any experiment. If someone does assume that the effect of temperature doesn't change when the quantity of the catalyst is changed, he or she will be incorrect. Following is a discussion of interaction and the practical way it should be handled.

The influence of temperature A differs with various quantities of catalyst B; hence, the curves showing the effect of temperature must be plotted separately for each catalyst quantity. But after the comparison of temperature and quantity of catalyst is finished, do we then have to investigate the effects of temperature plotted for various additives, types of catalyst, agitating conditions, the shapes of equipment, and so on?

Temperature and quantity of catalyst must be the only two factors among the many we intend to investigate. If so, we must consider the fact that the optimum conditions of temperature and quantity of the catalyst will differ according to different conditions of other factors, such as type of additive, quantity of additive, and others. In other words, there may be interactions between A, B, and other factors: C, D, ... The optimum conditions of A and B obtained from the curves of A and B may not be useful, since it may change when the type of additives changes.

This leads us to feel the need to plot the curves of temperature against all combinations of B, C, ... For this purpose we have to design a type of experiment to obtain the interactions between A and all the other factors. Hence, an experiment with only one factor, only two factors, or several factors would tend to be meaningless. Therefore, it would always be necessary to investigate the effects of all factors. If there are 10 factors on three levels, there are 59,049 different conditions.

If a research worker tried to obtain 59,049 curves of temperature A under different conditions of other factors in order to prepare a perfect report, this report could be 10,000 pages long. Clearly, no one can afford the time to observe 59,049 curves. We just want to know the range where the optimum temperature exists. Instead of getting 59,049 curves, the researchers should have conducted the next investigation.

Strictly speaking, the curves of temperature do differ under the conditions of other factors B, C, D, ... , but we must consider the average curve to be the representative curve of temperature A, and term it the *main effect* curve. Although curves do differ under the conditions of other factors; if the differences are not great, only a slight error will be caused when the optimum condition is determined from the main effect curve, which represents the effect of temperature A.

However, it is important to evaluate the extent of deviation from the main effect curve or the magnitude of the interaction. For this reason, the determination of interaction must be limited to the factors of concern to a research worker. It would be foolish to try to design an experiment for all interactions. It must be stressed that many interactions are not required; it is desirable that the magnitude of deviation be evaluated by the residual sum of squares.

When datum y_{ij} (the result of A_iB_j) is expressed by the composite effects of temperature A and quantity of catalyst B, we must find the extent of the error in such an expression. For this purpose we are going to determine the residual sum of squares, which includes interactions.

Since main effect A means the average curve of the effect of A, the means of the effects of A under conditions B_1, B_2, B_3, and B_4 are

$$\overline{A}_1 = \frac{A_1}{4} = \frac{-51}{4} = -\ 12.75$$

$$\overline{A}_2 = \frac{A_2}{4} = \frac{1}{4} = 0.25$$

$$\overline{A}_3 = \frac{A_3}{4} = \frac{15}{4} = 3.75$$

$$\overline{A}_4 = \frac{A_4}{4} = \frac{15}{4} = 3.75$$

$$\overline{A}_5 = \frac{A_5}{4} = \frac{8}{4} = 2.00 \qquad (32.5)$$

where A_1, A_2, A_3, A_4, and A_5 indicate the totals of the results under conditions A_1, A_2, A_3, A_4, and A_5.

Figure 32.1 shows the figures obtained after adding back the working mean, 80. The line of $\overline{T}$, which is 79.4, shows the average value of all experiments. Let a_1 be the difference between $\overline{A}_1$ and $\overline{T}$:

$$\hat{a}_1 = \frac{A_i}{4} - \frac{T}{20} \qquad (i = 1, 2, 3, 4, 5) \qquad (32.6)$$

If all five points were on-line $\overline{T}$, everyone would agree that within this range there is no effect due to changes in temperature on yield. It is true, therefore,

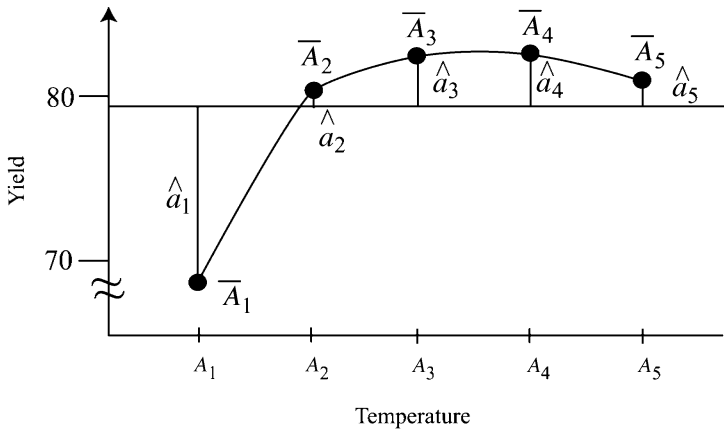

Figure 32.1
Effect of *A*

that the effect of temperature can be evaluated by making the following calculation:

$$\text{effect of } A = (\hat{a}_1)^2 + (\hat{a}_2)^2 + (\hat{a}_3)^2 + (\hat{a}_4)^2 + (\hat{a}_5)^2 \tag{32.7}$$

This is known as the *effect of the average curve.* Since there are four repetitions for each temperature, we can determine that four times the magnitude of equation (32.7) is the effect of temperature A, which constitutes part of the total variation of the data. Therefore, S_A, the effect of temperature A, is

$$\begin{aligned}
S_A &= 4[(\hat{a}_1)^2 + (\hat{a}_2)^2 + (\hat{a}_3)^2 + (\hat{a}_4)^2 + (\hat{a}_5)^2] \\
&= 4\left[\left(\frac{A_1}{4} - \frac{T}{20}\right)^2 + \left(\frac{A_2}{4} - \frac{T}{20}\right)^2 + \cdots + \left(\frac{A_5}{4} - \frac{T}{20}\right)^2\right] \\
&= 4\left[\frac{A_1^2}{16} - 2\,\frac{T}{20}\,\frac{A_1}{4} + \left(-\frac{T}{20}\right)^2\right] + \left[\frac{A_2^2}{16} - 2\,\frac{T}{20}\,\frac{A_2}{4} + \left(-\frac{T}{20}\right)^2\right] \\
&\quad + \cdots + \left[\frac{A_5^2}{16} - 2\,\frac{T}{20}\,\frac{A_5}{4} + -\frac{T^2}{20}\right] \\
&= \frac{A_1^2 + A_2^2 + A_3^2 + A_4^2 + A_5^2}{4} \\
&\quad - 2\,\frac{T}{20}\,\frac{A_1 + A_2 + A_3 + A_4 + A_5}{4}(4) + (4)(5)\left(\frac{T}{20}\right)^2 \qquad (32.8) \\
&= \frac{A_1^2 + A_2^2 + A_3^2 + A_4^2 + A_5^2}{4} - 2\,\frac{T^2}{20} + \frac{T^2}{20} \\
&= \frac{A_1^2 + A_2^2 + A_3^2 + A_4^2 + A_5^2}{4} - \frac{T^2}{20} \qquad (32.9)
\end{aligned}$$

To simplify equation (32.9), a working mean may be subtracted:

$$S_A = \frac{(-51)^2 + 1^2 + 15^2 + 15^2 + 8^2}{4} - \frac{(-12)^2}{20} = 772 \qquad (32.10)$$

In the calculation of S_A, $(A_1^2 + A_2^2 + \cdots + A_5^2)$ is divided by 4, since each of A_1, A_2, ... , A_5 is the total of four curves. Generally, in such calculations of variation, each squared value is divided by its *number of units*, which is the total of the square for each coefficient of a linear equation. Since A_1 is the sum of four observational values, A_1 is expressed by a linear equation with four observational values, where each value has a coefficient of 1. Therefore, A_1^2 must be divided by 4, the number of units of the linear equation, or the total of squares of each coefficient, 1. Similarly,

$$\begin{aligned} S_B &= \frac{B_1^2 + B_2^2 + B_3^2 + B_4^2}{5} - \text{CF} \\ &= \frac{(-44)^2 + (-9)^2 + 13^2 + 28^2}{5} - 7 = 587 \qquad (f = 3) \qquad (32.11) \end{aligned}$$

All other factor-related effects (including interaction $A \times B$) are evaluated by subtracting S_A and S_B from S_T, the total variations of all data. The residual sum of the squares, S_e, is called the *error variation:*

$$\begin{aligned} S_e &= S_T - S_A - S_B = 1593 - 772 - 587 \\ &= 234 \qquad (f = 19 - 4 - 3 = 12) \qquad (32.12) \end{aligned}$$

There are four unknown items in S_A; its degrees of freedom is four. If all the items are negligible, or if the effect of temperature A is negligible, S_A will simply be about equal to $4V_e$.

There are 4 units of error variance in S_A. Therefore, the net effect of A, or the pure variation S_A, is estimated by

$$S'_A = S_A - 4V_e \qquad (32.13)$$

Therefore,

$$\begin{aligned} S'_A &= 772 - (4)(19.5) \\ &= 694.0 \qquad (32.14) \end{aligned}$$

Similarly, the pure variation of B is

$$\begin{aligned} S'_B &= S_B - 3V_e = 587 - (3)(19.5) \\ &= 528.5 \qquad (32.15) \end{aligned}$$

The total effect of all factors, including A and B but excluding the general mean, is shown by S_T, which is 1593. The magnitude of all factor-related effects, excluding A and B, is then

$$\begin{aligned} S'_e &= (\text{total variation}) - (\text{pure variation of } A) - (\text{pure variation of } B) \\ &= 1593 - 694.0 - 528.5 = 370.5 \qquad (32.16) \end{aligned}$$

The yield results varied from 64 to 92%. The variation was caused by the effects of temperature, catalyst quantity, and other factors. How much of the total variation S_T was caused by temperature? This is calculated by dividing pure variation of A by S_T. If the percentage of variation caused by temperature is denoted by ρ_A, then

$$\rho_A = \frac{\text{pure variation of } A}{\text{total variation}} = \frac{S_A - 4V_e}{S_T}$$

$$= \frac{694}{1593} = 0.436 = 43.6\% \qquad (32.17)$$

That is, of the yield variation ranging from 64 to 92%, in 43.5% of the cases, the change of temperature caused the variation in yield. Similarly,

$$\rho_B = \frac{528.5}{1593} = 33.2\% \qquad (32.18)$$

$$\rho_e = \frac{370.5}{1593} = 23.2\% \qquad (32.19)$$

There are similarities between the analysis of variance and spectrum analysis. The total amount of variance, S_T, corresponds to the total power of the wave in spectrum analysis. In the latter, total power is decomposed to the power of each component with different cycles. Similarly, in analysis of variance, total variance is decomposed to a cause system. The ratio of the power of each component to total power is determined as degrees of contribution and is denoted by a percentage. The "components" for the purpose of our investigation are factor A, factor B, and other factors that change without artificial control. In the design of experiments, the cause system is designed to be orthogonal. Accordingly, calculation of the analysis of variance is easy and straightforward (Table 32.3).

Degrees of contribution can be explained as follows: In the case of a wave, when the amplitude of oscillation doubles, its power increases four fold. When the power is reduced to half, its amplitude is $1/\sqrt{2}$. In Table 32.1, the minimum value is 64 and the maximum value is 92.

We found from the analysis above that temperature and the quantity of the catalyst do indeed influence yield. What we want to know next is the relationship between temperature and yield and the quantity of catalyst and yield.

Table 32.3
Analysis of variance

Source	f	S	V	$E(V)$	S'	ρ (%)
A	4	772	193	$\sigma^2 + 4\sigma_A^2$	694.0	43.6
B	3	587	196	$\sigma^2 + 5\sigma_B^2$	528.5	33.2
e	12	234	19.5	σ^2	370.5	23.2
Total	19	1593			1593	100.0

In the previous calculation we had four estimates of yield at each temperature (200, 225, 250, 275, and 300°C). A total estimate for each temperature results in −51, 1, 15, 15, and 8, respectively. These totals are divided by 4, and the working mean 80 is added.

$$
\begin{aligned}
200°\text{C:} \quad & \frac{-51}{4} + 80 = 67.25 \\
225°\text{C:} \quad & \frac{1}{4} + 80 = 80.25 \\
250°\text{C:} \quad & \frac{15}{4} + 80 = 83.75 \\
275°\text{C:} \quad & \frac{8}{4} + 80 = 82.00
\end{aligned} \tag{32.20}
$$

Plotting these four points, a curve showing the relation of temperature and yield is obtained as shown in Figure 32.2. It is important to stress here that these values show the main effect of temperature and yield. The main effect curve shown in Figure 32.2 is the curve of the average values at B_1, B_2, B_3, and B_4.

Next, we want to find the temperature and quantity of catalyst that will contribute to an increase in yield. This procedure is actually the determination of optimum conditions. These conditions are normally found after observing the graphs in Figures 32.2 and Figure 32.3. The best temperature is the highest point of the curve. But the best temperature must result in the highest profit, and that may not necessarily mean the highest yield.

Actually, the heavy oil consumptions that may be necessary to maintain different temperatures are investigated first; these consumptions are then converted to the same cost per unit as that of the yield, and those units are then plotted against temperature. The temperature of the largest difference between yield and fuel cost is read from the abscissa, and it is then established as the operating standard.

From Figure 32.2 it is clear that the maximum value probably lies in the neighborhood of 250 to 275°C. If the value does not vary significantly, the lower temperature, 250°C, is the best temperature for our discussion. Similarly, the catalyst quantity that will give us the highest yield is determined from Figure 32.3. Let us assume that 0.8% is determined as the best quantity.

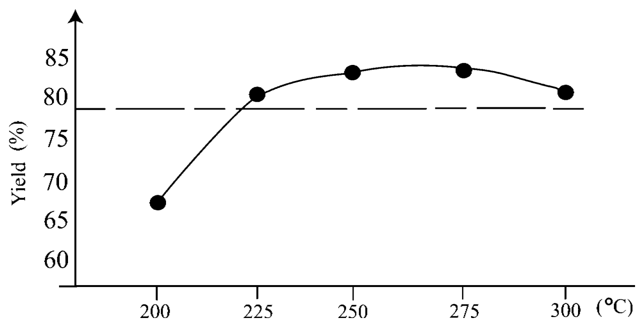

Figure 32.2
Relationship between temperature and yield

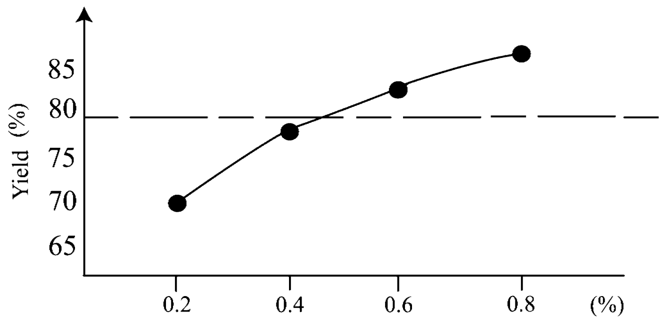

Figure 32.3
Relationship between catalyst and yield

Our next step is to find the actual value for the yield when production is continued at 250°C using an 0.8% catalyst quantity. To do this, we need to calculate the process average,expressed by the following equation, where $\hat{\mu}$ is the estimate of process average of yield at A_3B_4:

$$\begin{aligned}\hat{\mu} &= (\text{average of all experiments, } \overline{T})\\ &\quad + (\text{yield increase above } \overline{T} \text{ when } A_3 \text{ is used})\\ &\quad + (\text{yield increase above } \overline{T} \text{ when } B_4 \text{ is used})\\ &= \overline{T} + (\overline{A}_3 - \overline{T}) + (\overline{B}_4 - \overline{T}) = \overline{A}_3 + \overline{B}_4 - \overline{T}\\ &= 83.75 + 85.60 - 79.40 = 89.95 \qquad (32.21)\end{aligned}$$