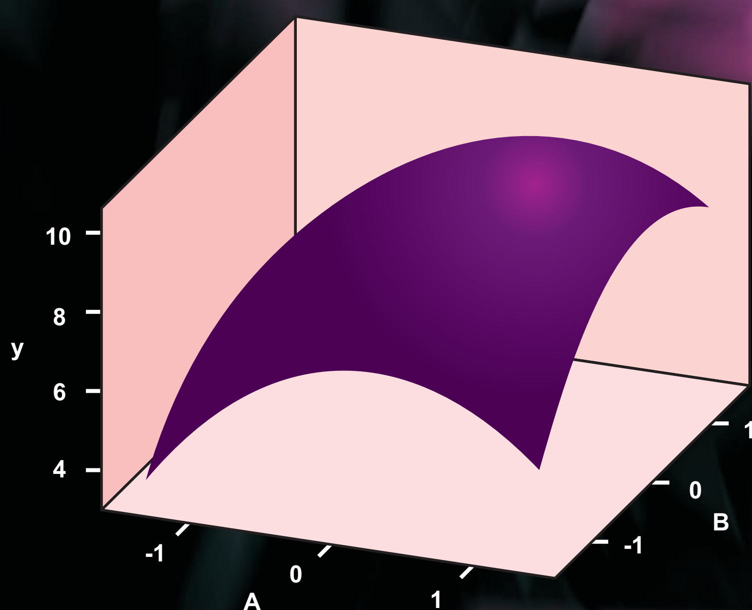
Eastern Economy Edition

# APPLIED DESIGN OF EXPERIMENTS AND TAGUCHI METHODS




K. Krishnaiah
P. Shahabudeen

# Applied Design of Experiments and Taguchi Methods

# Applied Design of Experiments and Taguchi Methods

**K. KRISHNAIAH**
Former Professor and Head
Department of Industrial Engineering
Anna University, Chennai

**P. SHAHABUDEEN**
Professor and Head
Department of Industrial Engineering
Anna University, Chennai

**PHI Learning Private Limited**
New Delhi-110001
2012

**₹ 395.00**

**APPLIED DESIGN OF EXPERIMENTS AND TAGUCHI METHODS**

K. Krishnaiah and P. Shahabudeen



**ISBN-978-81-203-4527-0**



Published by Asoke K. Ghosh, PHI Learning Private Limited, M-97, Connaught Circus, New Delhi-110001 and Printed by Baba Barkha Nath Printers, Bahadurgarh, Haryana-124507.

*To*

**My wife Kasthuri Bai and Students**

— K. Krishnaiah

*To*

**My Teachers and Students**

— P. Shahabudeen

# Contents

# Preface

Design of Experiments (DOE) is an off-line quality assurance technique used to achieve best performance of products and processes. This consists of (i) the design of experiment, (ii) conduct of experiment, and (iii) analysis of data. Designing the experiment suitable to a particular problem situation is an important issue in DOE. Robust design is a methodology used to design products and processes such that their performance is insensitive to noise factors. This book addresses the traditional experimental designs (Part I) as well as Taguchi Methods (Part II) including robust design. Though the subject of DOE is as old as Statistics, its application in industry is very much limited especially in the developing countries including India. One of the reasons could be that this subject is not taught in many academic institutions. However, this subject is being taught by the authors for the past fifteen years in the Department of Industrial Engineering, Anna University, Chennai. Dr. Krishnaiah has conducted several training programmes for researchers, scientists and industrial participants. He has also trained engineers and scientists in some organizations. Using their experience and expertise this book is written.

We hope that this book will be easy to follow on the first time reading itself. Those with a little or no statistical background can also find it very easy to understand and apply to practical situations. This book can be used as a textbook for undergraduate students of Industrial Engineering and postgraduate students of Mechanical Engineering, Manufacturing Systems Management, Systems Engineering and Operations Research, SQC & OR and Statistics.

We express our gratitude and appreciation to our colleagues in the Department of Industrial Engineering, Anna University, Chennai, for their encouragement and support. We also thank our faculty members Dr. M. Rajmohan, Dr. R. Baskaran and Mr. K. Padmanabhan and the research scholars Mr. S. Selvakumar and Mr. C. Theophilus for assisting in typing and correcting some of the chapters of the manuscript.

The case studies used in this book were conducted by our project/research students under our supervision. We are thankful to them. We thank our P.G. student Mr. L.M. Sathish for providing computer output for some of the examples. We also thank the students whose assignment problems/data have been used.

Our sincere thanks are due to the editorial and production teams of PHI Learning for their meticulous processing of the manuscript of this book.

Any suggestions for improvement of this book are most welcome.

**K. KRISHNAIAH**
**P. SHAHABUDEEN**

# PART I

# Design of Experiments

CHAPTER 1

# Review of Statistics

## 1.1 INTRODUCTION

**Population:** A large group of data or a large number of measurements is called *population*.

**Sample:** A sub set of data taken from some large population or process is a *sample*.

**Random sample:** If each item in the population has an equal opportunity of being selected, it is called a *random sample*. This definition is applicable for both infinite and finite population. A random sample of size $n$ if selected will be independently and identically distributed.

### *Sample statistics*

Suppose we have $n$ individual data ($X_1$, $X_2$, ..., $X_n$)

$$\text{Sample mean } (\bar{X}) = \frac{\sum X_i}{n}, \quad i = 1, 2, 3, ..., n \tag{1.1}$$

$$\text{Sample variance } (S^2) = \frac{\sum (X_i - \bar{X})^2}{n - 1} = \frac{\sum X_i^2 - n\bar{X}^2}{n - 1} \tag{1.2}$$

$$\text{Sample standard deviation } (S) = \sqrt{S^2} \tag{1.3}$$

### *Population parameters*

The population parameters for mean and standard deviation are denoted by $\mu$ and $\sigma$ respectively. The value of population parameter is always constant. That is, for any population data set, there is only one value of $\mu$ and $\sigma$.

## 1.2 NORMAL DISTRIBUTION

The normal distribution is a continuous probability distribution. It is a distribution of continuous random variables describing height of students, weight of people, marks obtained by students, process output measurements, etc. Usually the shape of normal distribution curve is bell shaped.

***Characteristics of normal distribution***

1. The curve is symmetric about the mean.
2. The total area under the curve is 1.0 or 100%.
3. The tail on either side of the curve extends to infinity.
4. The distribution is defined by two parameters $\mu$ and $\sigma$.

Though the curve extends to infinity, the curve does not touch the axis. Since the area beyond $\mu \pm 3\sigma$ is very small, for general use, we consider this area as zero. Knowing the two parameters of the curve, we can compute the area for any interval. Figure 1.1 shows a typical normal curve.

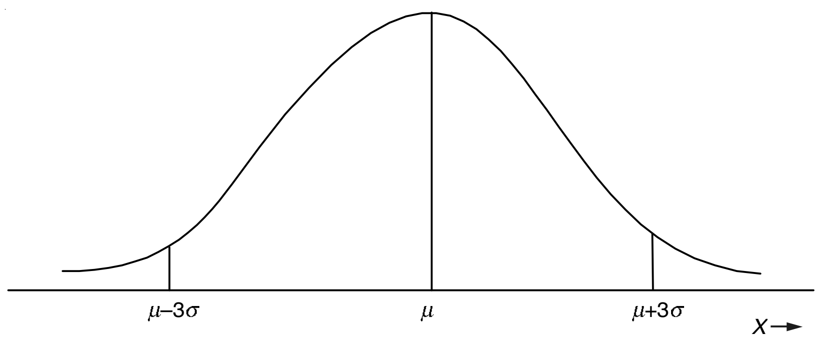



**FIGURE 1.1** Normal curve.

The probability density function for a normal distribution, $f(x)$ is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < \infty \tag{1.4}$$

where,

$e$ = 2.71828 and
$\pi$ = 3.14159 approximately.

$f(x)$ gives the vertical distance between the horizontal axis and the curve at point $x$.

If any data ($x$) is normally distributed, usually we represent it as $x \sim N(\mu, \sigma^2)$, indicating that the data is normally distributed with mean $\mu$ and variance $\sigma^2$.

### 1.2.1 Standard Normal Distribution

Standard normal distribution is a special case of normal distribution. This distribution facilitates easy calculation of area between any two points under the curve. Its mean is 0 and variance is 1. Suppose $x$ is a continuous random variable that has a normal distribution $N(\mu, \sigma^2)$, then the random variable

$$z = \frac{X - \mu}{\sigma}$$

follow standard normal distribution (Figure 1.2), denoted by $z \sim N(0, 1)$. The horizontal axis of this curve is represented by $z$. The centre point (mean) is labelled as 0. The $z$ values on the right side of the mean are positive and on the left side are negative. The $z$ value for a point ($x$) on the horizontal axis gives the distance between the mean and that point in terms of the standard deviation. For example, a point with a value of $z = 1$, indicates that the point is 1 standard deviation to the right of the mean. And $z = -2$ indicates that the point is 2 standard deviations to the left of the mean. Figure 1.2 shows the standard normal distribution. The standard normal distribution table is given in Appendix A.1. This table gives the areas under the standard normal curve between $z = 0$ and the values of $z$ from 0.00 to 3.09. Since the total area under the curve is 1.00 and the curve is symmetric about the mean, the area on each side of the mean is 0.5.

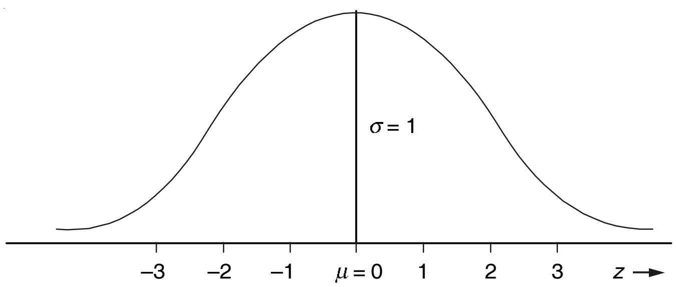



**FIGURE 1.2** Standard normal distribution curve.

For computing the area under the curve from $z = 0$ ($-\infty$) and any point $x$, we compute the value of

$$z_x = \frac{x - \mu}{\sigma} \tag{1.5}$$

Corresponding to this $z$ value, we obtain the area from Table A.1.

## ILLUSTRATION 1.1

The diameter of shafts manufactured is normally distributed with a mean of 3.0 cm and a standard deviation of 0.009 cm. The shafts that are with 2.98 cm or less diameter are scrapped and shafts with diameter more than 3.02 cm are reworked. Determine the percentage of shafts scrapped and percentage of rework.

**SOLUTION:**

Mean ($\mu$) = 3.0 cm
Standard deviation ($\sigma$) = 0.009 cm
Let upper limit for rework ($U$) = 3.02 cm
Lower limit at which shafts are scrapped ($L$) = 2.98
Now let us determine the $Z$ value corresponding to $U$ and $L$

$$Z_U = \frac{U - \mu}{\sigma} = \frac{3.02 - 3.00}{0.009} = 2.22$$

$$Z_L = \frac{L - \mu}{\sigma} = \frac{2.98 - 3.00}{0.009} = -2.22$$

From standard normal tables $P(Z_U > 2.22) = 0.5 - 0.4868 = 0.0132$ or 1.32%
That is, percentage of rework = 1.32

Similarly, $P(Z_L < -2.22) = 0.5 - 0.4868 = 0.0132$ or 1.32% (scrap)

Figure 1.3 shows the probability calculation for the Illustration 1.1.

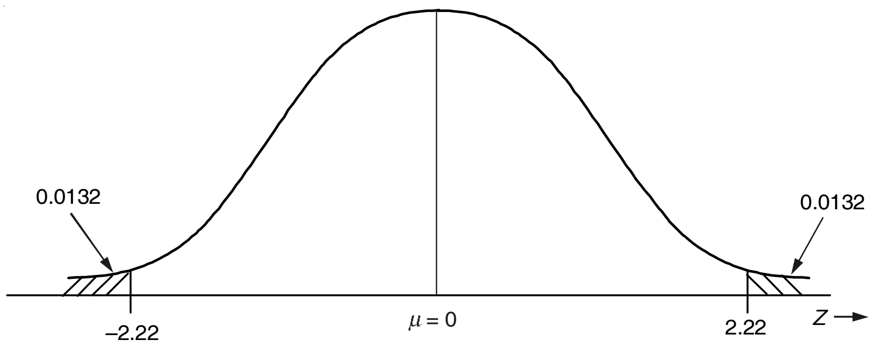



**FIGURE 1.3** Finding tail area beyond $Z$ value.

## 1.3. DISTRIBUTION OF SAMPLE MEANS

The distribution of a sample statistic is called sampling distribution. Suppose we draw $m$ samples of size $n$ from a population. The value of each sample mean ($\bar{X}$) will be different and the sample mean $\bar{X}$ is a random variable. The distribution of these sample means is termed sampling distribution of $\bar{X}$.

### *Mean of the sampling distribution*

The mean of the sampling distribution is the mean of all the sample means and is denoted by $\mu_{\bar{X}}$. The mean of the population ($\mu$) is estimated by $\mu_{\bar{X}}$.

### *Standard deviation of the sampling distribution*

The standard deviation of the sampling distribution is denoted by $\sigma_{\bar{X}}$ and is equal to $\sigma/\sqrt{n}$. That is

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \tag{1.6}$$

Equation 1.6 is applicable when $n/N \leq 0.05$. Else, we have to use a correction factor $\sqrt{\dfrac{N-n}{N-1}}$. That is

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \tag{1.7}$$

The standard deviation of sampling distribution $\sigma_{\bar{X}}$ is also called *standard error of mean*.

***Shape of sampling distribution***

The shape of the sampling distribution depends on whether the samples are drawn from normal population or non-normal population. If the samples are drawn from normal population $N(\mu, \sigma^2)$, the shape of its sampling distribution is also normal. If samples are drawn from non-normal population, the shape of its sampling distribution will be approximately normal (from central limit theorem). As the sample size increases ($n \geq 30$), the shape of sampling distribution is approximately normal irrespective of the population distribution. So, in general we can make use of the characteristics of normal distribution for studying the distribution of sample means.

Note that $N(\mu, \sigma^2)$, indicates normal population with mean $\mu$ and variance $\sigma^2$.

## 1.4 THE *t*-DISTRIBUTION

The *t*-distribution is also known as student's *t*-distribution. It is similar to normal distribution in some aspects. The *t*-distribution is also symmetric about the mean. It is some what flatter than the normal curve. As the sample size increases, the *t*-distribution approaches the normal distribution. The shape of the *t*-distribution curve depends on the number of degrees of freedom. The degrees of freedom for *t*-distribution are the sample size minus one. The standard deviation of *t*-distribution is always greater than one. The *t*-distribution has only one parameter, the degrees of freedom.

Suppose $X_1, X_2, ..., X_n$ is a random sample from $N(\mu, \sigma^2)$ distribution. If $\bar{X}$ and $S^2$ are computed from this sample are independent, the random variable $\dfrac{\bar{X} - \mu}{S/\sqrt{n}}$ has a *t*-distribution with $n - 1$ degrees of freedom.

Table of percentage points of the *t*-distribution is given in Appendix A.2. Its application is discussed in Section 1.7.

## 1.5 THE *F*-DISTRIBUTION

The *F*-distribution is defined by two numbers of degrees of freedom, the numerator degrees of freedom ($v_1$) and the denominator degrees of freedom ($v_2$). The distribution is skewed to right and decreases with increase of degrees of freedom. The *F*-statistic is named after Sir Ronald Fisher. The *F*-statistic is used to test the hypothesis in ANOVA. Table of percentage points of the *F*-distribution is given in Appendix A.3. The value in *F*-table gives the right tail area for a given set of $v_1$ and $v_2$ degrees of freedom.

## 1.6 CONFIDENCE INTERVALS

We often estimate the value of a parameter, say the height of college's male students from a random sample of size *n* by computing the sample mean. This sample mean is used to estimate the population mean. Such an estimate is called *point estimate*. The accuracy of this estimate largely depends on the sample size. It always differs from the true value of population mean.

Instead, we use an interval estimate by constructing around the point estimate and we make a probability statement that this interval contains the corresponding population parameter. These interval statements are called *confidence intervals*. The extent of confidence we have that this interval contains the true population parameter is called the *confidence level*. It is denoted by $(1-\alpha)$ 100%, where $\alpha$ is called the *significance level*. And $(1-\alpha)$ is called the *confidence coefficient*.

### 1.6.1 Confidence Interval on a Single Mean ($\mu$)

**Case 1: Large samples** ($n > 30$)

The $(1-\alpha)$ 100% confidence interval for $\mu$ is

$$\bar{X} \pm Z_{\alpha/2}\,\sigma_{\bar{X}}, \text{ if } \sigma \text{ is known} \tag{1.8}$$

$$\bar{X} \pm Z_{\alpha/2}\,S_{\bar{X}}, \text{ if } \sigma \text{ is not known} \tag{1.9}$$

where, $Z$ is the standard normal deviate corresponding to the given confidence level.

**Case 2: Small samples** ($n < 30$)

The $(1-\alpha)$ 100% confidence interval for $\mu$ is

$$\bar{X} \pm t_{\alpha/2,\,n-1}\,S_{\bar{X}} \tag{1.10}$$

where the value of $t$ is obtained from the $t$-distribution corresponding to $n-1$ degrees of freedom for the given confidence level.

### 1.6.2 Confidence Interval for Difference in Two Means

**Case 1: Large samples**

The $(1-\alpha)$ 100% confidence interval for $\mu_1 - \mu_2$ is

$$\bar{X}_1 - \bar{X}_2 \pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \text{ if } \sigma_1 \text{ and } \sigma_2 \text{ are known} \tag{1.11}$$

$$\bar{X}_1 - \bar{X}_2 \pm Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}, \text{ if } \sigma_1 \text{ and } \sigma_2 \text{ are not known} \tag{1.12}$$

**Case 2: Small samples**

The $(1-\alpha)$ 100% confidence interval for $\mu_1 - \mu_2$ is

$$\bar{X}_1 - \bar{X}_2 \pm t_{\alpha/2}\,S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \tag{1.13}$$

where
$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

The $t$-value is obtained from $t$-distribution for the given confidence level and $n_1 + n_2 - 2$ degrees of freedom.

## 1.7 HYPOTHESES TESTING

A statistical hypothesis is an assumption about the population being sampled. There are two types of hypothesis.

1. Null hypothesis ($H_0$)
2. Alternative hypothesis ($H_1$)

### *Null hypothesis*

A null hypothesis is a claim or statement about a population parameter, that is assumed to be true. For example, a company manufacturing electric bulbs claims that the average life of their bulbs ($\mu$) is 1000 hours. In reality it may or may not be true. If it is true, the average life $\mu = 1000$ hr. For this, the null hypothesis is written as

$$H_0\text{: } \mu = 1000 \text{ hr}$$

### *Alternative hypothesis*

An alternative hypothesis is a claim or a statement about a population parameter, that is true if null hypothesis is false. For example, the alternative hypothesis of life of bulbs, is that the average life of bulbs is less than 1000 hr. It is written as

$$H_1\text{: } \mu < 1000 \text{ hr}$$

A test of hypothesis is simply a rule by which a hypothesis is either accepted or rejected. Such a rule is usually based on sample statistics called *test statistics*. Since it is based on sample statistics computed from $n$ observations, the decision is subject to two types of errors.

**Type I error:** The null hypothesis is true, but rejected. The probability of Type I error is denoted by $\alpha$. The value of $\alpha$ represents the significance level of the test.

$$\alpha = P(H_0 \text{ is rejected} \mid H_0 \text{ is true})$$

**Type II error:** The hypothesis is accepted when it is not true. That is, some alternative hypothesis is true. The probability of Type II error is denoted by $\beta$.

$$\beta = P(H_0 \text{ is accepted} \mid H_0 \text{ is false})$$

$(1 - \beta)$ is called the *power of the test*. It denotes the probability of not committing the Type II error.

***Tails of a test:*** Depending on the type of alternative hypothesis, we have either one tail test or two tail test. Suppose we have the null and alternative hypothesis as follows:

$$H_0\text{: } \mu = \mu_0$$

$$H_1\text{: } \mu \neq \mu_0$$

In this case, we have a two tail test (Figure 1.4). In a two tail test the rejection region will be on both tails and the $\alpha$ value is equally divided.

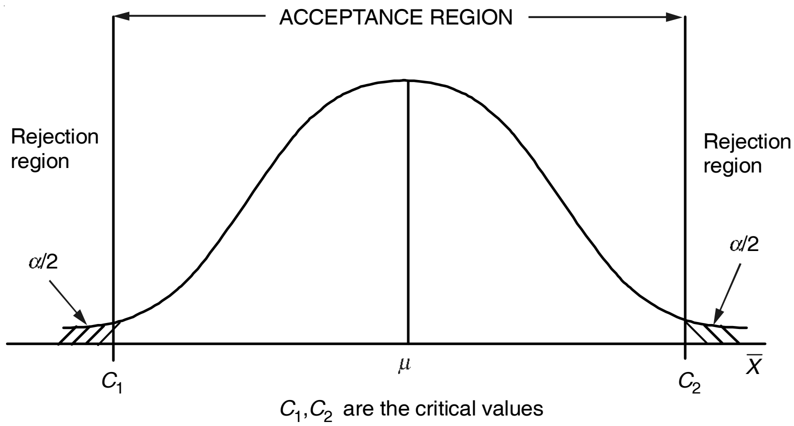



**FIGURE 1.4** A two tailed test.

In the case of one tail test, the rejection region will exist only on one side of the tail depending on the type of alternative hypothesis. If the alternative hypothesis is

$H_1$: $\mu > \mu_0$, it is a right tail test, and if
$H_1$: $\mu < \mu_0$, it is a left tail test.

### *p-value approach*

*P*-value approach is defined as the smallest value of significance level ($\alpha$) at which the stated null hypothesis is rejected. Here we try to determine the *p*-value for the test. If we have a predetermined value of $\alpha$, we compare the *p*-value with $\alpha$ and arrive at a decision.

We reject the null hypothesis if

$$\alpha > p\text{-value} \quad \text{or} \quad p\text{-value} < \alpha$$

and we do not reject the null hypothesis if

$$\alpha \le p\text{-value} \quad \text{or} \quad p\text{-value} \ge \alpha$$

For a one tail test, the *p*-value is given by the area in the tail of the sampling distribution curve beyond the observed value of the sample statistic. Figure 1.5 shows the *p*-value for a left tail test. For a two tail test, the *p*-value is twice the area in the tail of the sampling distribution curve beyond the observed sample statistic. Figure 1.6 shows the *p*-value for a two tail test.

Suppose we compute the value of $Z$ for a test as $Z = (\bar{X} - \mu)/\sigma_{\bar{x}}$. This value is called *observed value* of $Z$. Then we find the area under the tail of the normal distribution curve beyond this value of $Z$. This area gives the *p*-value or one-half *p*-value depending on whether it is a one tail test or two tail test.

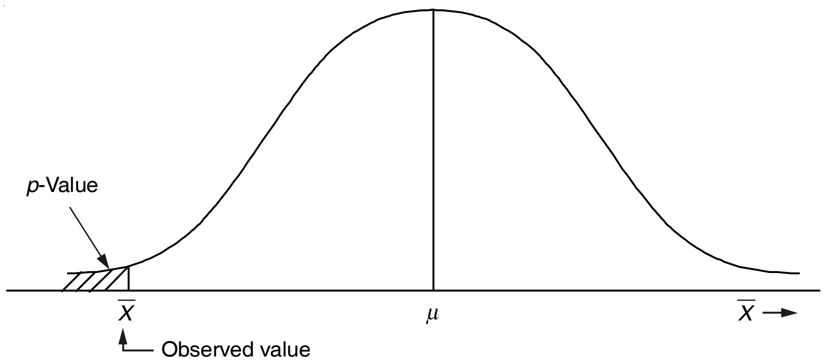



**FIGURE 1.5** *p*-value for a left tailed test.

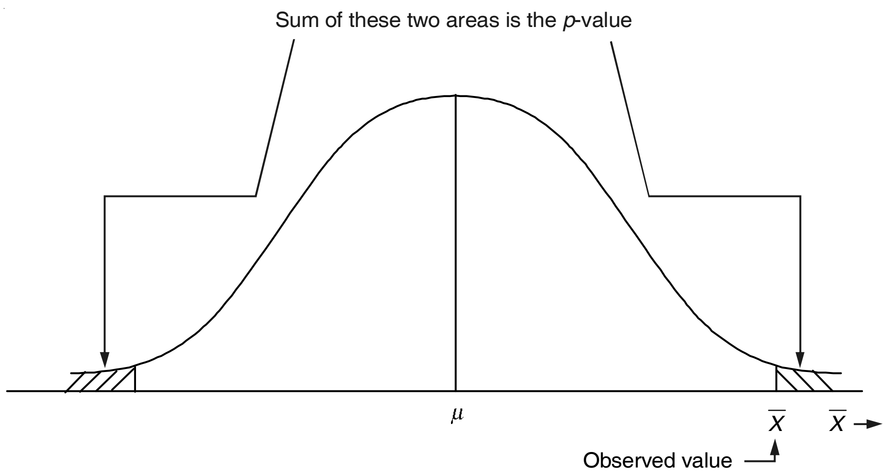



**FIGURE 1.6** *p*-value for a two tailed test.

### 1.7.1 Tests on a Single Mean

The null and alternative hypothesis for this test is as follows:

$H_0$: $\mu = \mu_0$
$H_1$: $\mu \neq \mu_0$ or $\mu > \mu_0$ or $\mu < \mu_0$ (depends on the type of problem)

**Case 1: When $\sigma$ is known**
The test statistic is

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} \tag{1.14}$$

where $\overline{X}$ is the sample mean and $n$ is the sample size and $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$

Reject $H_0$ if $|Z_0| > Z_{\alpha/2}$ (or $Z_0 > Z_\alpha$ or $Z_0 < -Z_\alpha$ depending on type of $H_1$)

This test is applicable for a normal population with known variance or if the population is non-normal but the sample size is large ($n > 30$), in which case $\sigma_{\bar{X}}$ is replaced by $S_{\bar{X}}$ in Eq. (1.14).

## ILLUSTRATION 1.2

The tensile strength of fabric is required to be at least 50 kg/cm$^2$. From past experience it is known that the standard deviation of tensile strength is 2.5 kg/cm$^2$. From a random sample of 9 specimens, it is found that the mean tensile strength is 48 kg/cm$^2$.

(i) State the appropriate hypotheses for this experiment and test the hypotheses using $\alpha = 0.05$. What is your conclusion?
(ii) What is your decision based on the $p$-value?

**SOLUTION:**

(i) The hypotheses to be tested are

$H_0$: $\mu \geq 50$ kg/cm$^2$
$H_1$: $\mu < 50$ kg/cm$^2$

Since the standard deviation is known, the test statistic is

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma_{\bar{x}}} = \frac{48 - 50}{2.5/\sqrt{9}}$$

$$= \frac{-2}{0.83} = -2.41$$

Reject $H_0$, if $Z_0 < -Z_{0.05}$
From standard normal table, the critical value for $Z_{0.05} = 1.65$.
Hence, we reject the null hypothesis and conclude that the tensile strength is less than 50 kg/cm$^2$.

(ii) $p$-value approach: From standard normal table, the tail area under the curve beyond –2.41 is 0.008.

So, the $p$-value for the test is 0.008.
Since the $\alpha$-value is more than the $p$-value, we reject the null hypothesis.

## ILLUSTRATION 1.3

A study was conducted a year back which claims that the high school students spend on an average 11 hours per week on Internet. From a sample of 100 students studied recently found that they spend on average 9 hours per week on Internet with a standard deviation of 2.2 hours.

(i) Test the hypotheses that the current students spend less than 11 hours on Internet. Use $\alpha = 0.05$.
(ii) What is the $p$-value for the test?
(iii) Determine the 95% confidence interval for the mean time.

**SOLUTION:**

(i) Note that the sample size is large ($n > 30$) and hence $Z$ statistic is applicable.

$H_0$: $\mu = 11$ hr
$H_1$: $\mu < 11$ hr

$$Z_0 = \frac{\bar{X} - \mu_0}{S_{\bar{x}}} = \frac{9 - 11}{2.2/\sqrt{100}} = -\frac{2}{0.22} = -9.09$$

Reject $H_0$, if $Z_0 < -Z_{0.05}$
From standard normal table, the critical value for $Z_{0.05} = 1.65$.

Therefore, we reject the null hypothesis. The conclusion is that the current students on average spend less time than the time found earlier.

(ii) From standard normal tables, corresponding to $Z = -9.09$, the area under the curve can be taken as zero.
Hence, the $p$-value = 0.
Since $\alpha = 0.05 > p$-value, we reject the null hypothesis.

(iii) The confidence interval is given by

$$\bar{X} \pm Z_{\alpha/2}\,\sigma_{\bar{X}} = 9 \pm 1.96\left(\frac{2.2}{\sqrt{100}}\right)$$
$$= 9 \pm 0.43$$
$$8.57 \le \mu \le 9.43$$

**Case 2: When $\sigma$ is unknown normal population**

The test statistic is

$$t_0 = \frac{\bar{X} - \mu_0}{S_{\bar{x}}} \quad (\text{when } n < 30) \tag{1.15}$$

where, $S_{\bar{x}} = S/\sqrt{n}$

Reject $H_0$ if $|t_0| > t_{\alpha/2, n-1}$ or $t_0 < -t_{\alpha, n-1}$ or $t_0 > t_{\alpha, n-1}$ (depending on type of $H_1$)

## ILLUSTRATION 1.4

A cement manufacturer claims that the mean settling time of his cement is not more than 45 minutes. A random sample of 20 bags of cement selected and tested showed an average settling time of 49.5 minutes with a standard deviation of 3 minutes.

Test whether the company's claim is true. Use $\alpha = 0.05$.

**SOLUTION:**

Here the sample size is small ($n < 30$). Hence, we use the $t$-statistic.

$H_0$: $\mu \le 45$ minutes
$H_1$: $\mu > 45$ minutes

$$t_0 = \frac{\bar{X} - \mu_0}{S_{\bar{x}}} = \frac{49.5 - 45}{3/\sqrt{20}} = \frac{4.5}{0.67} = 6.7$$

Reject $H_0$ if $t_0 > t_{\alpha, n-1}$

From $t$-table, $t_{0.05,19} = 1.729$. Hence we reject the null hypothesis. The inference is that the settling time is greater than 45 minutes.

**ILLUSTRATION 1.5**

A gym claims that their weight loss exercise causes an average weight reduction of at least 10 kg. From a random sample of 36 individuals it was found that the average weight loss was 9.5 kg with a standard deviation of 2.2 kg.

(i) Test the claim of the gym. Use $\alpha = 0.05$.
(ii) Find the $p$-value for the test.

**SOLUTION:**

(i) The null and alternative hypotheses are:

$H_0$: $\mu \geq 10$ kg
$H_1$: $\mu < 10$ kg

$$Z_0 = \frac{\overline{X} - \mu_0}{S_{\overline{x}}} = \frac{9.5 - 10}{2.2/\sqrt{36}} = -1.35$$

Reject $H_0$, if $Z_0 < -Z_{0.05}$
From standard normal table, the critical value for $Z_{0.05} = 1.65$.
Hence, we do not reject the null hypothesis. That is, the claim made by the gym is true.

(ii) Corresponding to the $Z$ value of 1.35, from standard normal tables, the probability is 0.0885.
That is, the $p$-value = 0.0885.
Since the $\alpha$-value is less than the $p$-value, we do not reject the null hypothesis.

### 1.7.2 Tests on Two Means

Here the two samples are assumed as independent. When we have two population means $\mu_1$ and $\mu_2$, we test the hypothesis $\mu_1 - \mu_2$. The alternative hypothesis may be

1. The two population means are different
   $\mu_1 \neq \mu_2$ which is same as $\mu_1 - \mu_2 \neq 0$
2. The mean of the first population is more than the second population mean
   $\mu_1 > \mu_2$ which is equivalent to $\mu_1 - \mu_2 > 0$
3. The mean of the first population is less than the second population mean
   $\mu_1 < \mu_2$ which is equivalent to $\mu_1 - \mu_2 < 0$

**Case 1: When variances are known $\sigma_1^2$ and $\sigma_2^2$**

$H_0$: $\mu_1 = \mu_2$
$H_1$: $\mu_1 \neq \mu_2$ or $\mu_1 > \mu_2$ or $\mu_1 < \mu_2$

The test statistic is

$$Z_0 = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \tag{1.16}$$

The value of $\mu_1 - \mu_2$ in Eq. (1.16) is substituted from $H_0$.

Reject $H_0$ if $|Z_0| > Z_{\alpha/2}$ (or $Z_0 > Z_\alpha$ or $Z_0 < -Z_\alpha$ depending on type of $H_1$)

If variances are not known and sample size is large ($n > 30$), $\sigma_1$ and $\sigma_2$ in Eq. (1.16) are replaced by $S_1$ and $S_2$.

## ILLUSTRATION 1.6

A company manufacturing clay bricks claims that their bricks (Brand *A*) dry faster than its rival company's Brand *B*. A customer tested both brands by selecting samples randomly and the following results have been obtained (Table 1.1).

**TABLE 1.1** Illustration 1.6

| *Brand* | *Sample size* | *Mean drying time* (hr) | *Standard deviation of drying time* (hr) |
|---|---|---|---|
| *A* | 25 | 44 | 11 |
| *B* | 22 | 49 | 9 |

Test whether the company's claim is true at 5% significance level. Also construct the 95% confidence interval for the difference in the two means.

**SOLUTION:**

$H_0$: $\mu_1 = \mu_2$
$H_1$: $\mu_1 < \mu_2$

The test statistic is

$$Z_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \qquad (\mu_1 - \mu_2 = 0 \text{ from } H_0)$$

Reject $H_0$ if $Z_0 < -Z_\alpha$

$$Z_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \frac{44 - 49}{\sqrt{\dfrac{11^2}{25} + \dfrac{9^2}{22}}} = -\frac{5}{2.92} = -1.71$$

From normal table, $Z_{0.05} = 1.65$. Hence we reject $H_0$. That is, the mean drying time of both the brands is different.

***p*-value approach:** The $p$-value for the test is 0.0436.

Since the $\alpha$-value is more than the $p$-value, we reject the null hypothesis.

**Confidence interval:** The $(1 - \alpha)$ 100% confidence interval for $\mu_1 - \mu_2$ is

$$\bar{X}_1 - \bar{X}_2 \pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 5 \pm 1.96(2.92) = 5 \pm 5.72$$

or
$$-0.72 \le \mu_1 - \mu_2 \le 10.72$$

**Case 2: When variances unknown: normal populations ($\sigma_1^2 = \sigma_2^2$)**

$H_0$: $\mu_1 = \mu_2$
$H_1$: $\mu_1 \neq \mu_2$

The test statistic is

$$t_0 = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{(n_1 - 1)\,S_1^2 + (n_2 - 1)\,S_2^2}{n_1 + n_2 - 2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \qquad (1.17)$$

The value of $\mu_1 - \mu_2$ is substituted from $H_0$.

Reject $H_0$ if $|t_0| > t_{\alpha/2,\,v}$ where $v = n_1 + n_2 - 2$.

Applicable when the sample sizes are less than 30.

## ILLUSTRATION 1.7

In the construction industry a study was undertaken to find out whether male workers are paid more than the female workers. From a sample of 25 male workers, it was found that their average wages was ₹ 115.70 with a standard deviation of ₹ 13.40. Whereas the average wages of female workers were ₹ 106.0 with a standard deviation of ₹ 10.20 from a sample of 20. Assume that the wages follow normal distribution with equal but unknown population standard deviations. Using 5% significance level, test whether the wages of male workers is same as that of female workers.

**Solution:**

Here it is assumed that the standard deviations are unknown but are equal.

$H_0$: $\mu_1 = \mu_2$
$H_1$: $\mu_1 \neq \mu_2$

The test statistic is

$$t_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{(n_1 - 1)\,S_1^2 + (n_2 - 1)\,S_2^2}{n_1 + n_2 - 2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

Reject $H_0$ if $|t_0| > t_{\alpha/2,\,v}$ where $v = n_1 + n_2 - 2$

$$t_0 = \frac{115.7 - 106.0}{\sqrt{\dfrac{24 \times 13.4^2 + 19 \times 10.2^2}{25 + 20 - 2}\left(\dfrac{1}{25} + \dfrac{1}{20}\right)}} = \frac{9.7}{3.63} = 2.672$$

From $t$-table, $t_{0.025,43} = 2.017$. Hence, we reject $H_0$. That is, the average wages paid to male and female workers is significantly different.

***The p-value approach:*** To find the $p$-value we first find the significance level corresponding to the tail area $t = 2.672$ at 43 degrees of freedom. From $t$-table we may not always be able to find the tail area matching to the computed value ($t_0$). In such a case we try to find the nearest area to $t_0$. Now for this problem the nearest tail area for $t_0 = 2.672$ at 43 degrees of freedom is 2.695 and the corresponding significance level is 0.005. Since it is a two tail test, the approximate $p$-value is equal to $0.005 \times 2 = 0.01$.

Since the $\alpha$-value is more than the $p$-value, we reject the null hypothesis.

**Case 3: When variances unknown: normal populations** $(\sigma_1^2 \neq \sigma_2^2)$

$H_0$: $\mu_1 = \mu_2$

$H_1$: $\mu_1 > \mu_2$ or $\mu_1 < \mu_2$

The test statistic is

$$t_0 = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \text{ with } \nu \text{ degrees of freedom} \tag{1.18}$$

where,

$$\nu = \frac{\left(\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right)^2}{\dfrac{(S_1^2/n_1)^2}{n_1 - 1} + \dfrac{(S_2^2/n_2)^2}{n_2 - 1}} \tag{1.19}$$

Reject $H_0$ if $|t_0| > t_{\alpha/2,\nu}$ or $t_0 > t_{\alpha,\nu}$ or $t_0 < -t_{\alpha,\nu}$

Applicable when the sample sizes are less than 30.

## ILLUSTRATION 1.8

A study on the pattern of spending in shopping has been conducted to find out whether the spending is same between male and female adult shoppers. The data obtained are given in Table 1.2.

**TABLE 1.2** Illustration 1.8

| *Population* | *Sample size* | *Average amount spent* (₹) | *Standard deviation* (₹) |
|---|---|---|---|
| Males | 25 | 80 | 17.5 |
| Females | 20 | 96 | 14.4 |

Assume that the two populations are normally distributed with unknown and unequal variances. Test whether the difference in the two means is significant at 5% level.

**SOLUTION:**

$H_0$: $\mu_1 - \mu_2 = 0$
$H_1$: $\mu_1 - \mu_2 \neq 0$

It is a two tail test. The area in each tail = $\alpha/2 = 0.025$.
The test statistic is

$$t_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} = \frac{80 - 96}{\sqrt{\dfrac{17.5^2}{25} + \dfrac{14.4^2}{20}}} = -\frac{16}{4.76} = -3.36$$

$$v = \frac{\left(\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right)^2}{\dfrac{(S_1^2/n_1)^2}{n_1 - 1} + \dfrac{(S_2^2/n_2)^2}{n_2 - 1}} = \frac{(12.25 + 10.37)^2}{\dfrac{150.06}{24} + \dfrac{107.5}{19}} = \frac{511.66}{11.91} = 43$$

Reject $H_0$ if $|t_0| > t_{\alpha/2,v}$

From $t$-table, $t_{\alpha/2,v} = t_{0.025,43} = 2.017$

Since $|t_0| > t_{\alpha/2,\ v}$ we reject $H_0$. That is, the difference between the two means is significant, which means that female shoppers spend more than male shoppers.

The approximate $p$-value for this test is $2 \times 0.001 = 0.002$.

### 1.7.3 Dependent or Correlated Samples

Here we have the same sample before and after some treatment (test) has been applied. Suppose we want to test the effect of a training program on a group of participants using some criteria. We evaluate the group before the training program and also after the training program using the same criteria and then statistically test the effect. That is, the same sample is being used before and after the treatment. Thus, we will have $n$ pairs of data. Such samples are called *dependant* or *correlated samples*. The test employed in such cases is called *paired* $t$-test. The procedure is to take the differences between the first and second observation on the same sample (person or part) and the mean difference ($\mu_d$) is tested. The hypotheses is

$H_0$: $\mu_d = 0$
$H_1$: $\mu_d \neq 0$

The test statistic is

$$t_0 = \frac{\bar{d}}{S_d/\sqrt{n}} \quad \text{with } n - 1 \text{ degrees of freedom} \tag{1.20}$$

where $\bar{d}$ is the average difference of $n$ pairs of data and $S_d$ is standard deviation of these differences.

Reject $H_0$ if $|t_0| > t_{\alpha/2,\, n-1}$

**ILLUSTRATION 1.9**

Two types of assembly fixtures have been developed for assembling a product. Ten assembly workers were selected randomly and were asked to use these two fixtures to assemble the products. The assembly time taken for each worker on these two fixtures for one product is given in Table 1.3.

**TABLE 1.3** Illustration 1.9

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Fixture 1 | 23 | 26 | 19 | 24 | 27 | 22 | 20 | 18 | 21 | 25 |
| Fixture 2 | 21 | 24 | 23 | 25 | 24 | 28 | 24 | 23 | 19 | 22 |

Test at 5% level of significance whether the mean times taken to assemble a product are different for the two types of fixtures.

**SOLUTION:**

Here, each worker assembles the same product using both the fixtures. Hence the samples are dependent. So, we have to use paired $t$ test.

$H_0$: $\mu_d = 0$
$H_1$: $\mu_d \neq 0$

The test statistic is (Eq. 1.20)

$$t_0 = \frac{\bar{d}}{S_d/\sqrt{n}} \quad \text{with } n-1 \text{ degrees of freedom}$$

Reject $H_0$ if $|t_0| > t_{\alpha/2,n-1}$

| Product | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Fixture 1 | 23 | 26 | 19 | 24 | 27 | 22 | 20 | 18 | 21 | 25 |
| Fixture 2 | 21 | 24 | 23 | 25 | 24 | 28 | 24 | 23 | 19 | 22 |
| Difference ($d$) | 2 | 2 | – 4 | –1 | 3 | – 6 | – 4 | –5 | 2 | 3 |
| $d^2$ | 4 | 4 | 16 | 1 | 9 | 36 | 16 | 25 | 4 | 9 |

The values of $\bar{d}$ and $S_d$ are computed as follows:

$\Sigma d = -8$, $\Sigma d^2 = 124$ and $n = 10$

$$\bar{d} = \frac{\sum d}{n} = -\frac{8}{10} = -0.80$$

$$S_d = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}} = \sqrt{\frac{124 - \frac{(-8)^2}{10}}{9}} = 3.61$$

$$t_0 = \frac{\bar{d}}{S_d/\sqrt{n}} = \frac{-0.80}{3.61/\sqrt{10}} = -\frac{0.8}{1.142} = -0.7$$

From table $t_{\alpha/2,n-1} = t_{0.025,9} = 2.262$

Since $|t_0| < t_{\alpha/2,n-1}$, we do not reject the null hypothesis. That is, there is no significant difference between the assembly times obtained by both fixtures. Hence any one fixture can be used.

## PROBLEMS

**1.1** A random sample of 16 observations taken from a population that is normally distributed produced a mean of 812 with a standard deviation of 25. Test the hypothesis that the true mean is 800.

**1.2** The breaking strength of clay bricks is required to be at least 90 kg/cm$^2$. From past experience it is known that the standard deviation of tensile strength is 9 kg/cm$^2$. From a random sample of 25 bricks, it is found that the mean breaking strength is 80 kg/cm$^2$.
(i) State the appropriate hypotheses for this experiment
(ii) Test the hypotheses. Use $\alpha = 0.05$. What is your decision?
(iii) Find the $p$-value for the test.

**1.3** The desired breaking strength of clay bricks is 85 kg/cm$^2$. From a random sample of 49 bricks it is found that the average breaking strength is 75 kg/cm$^2$ with a standard deviation of 5 kg/cm$^2$.
(i) State the appropriate hypotheses for this experiment and test it. Use $\alpha = 0.05$.
(ii) Verify your decision using $p$-value.

**1.4** The life of an electric bulb (hours) is of interest. Ten bulbs are selected randomly and tested and the following results are obtained.

850, 900, 690, 800, 950, 700, 890, 670, 800, 880

(i) Test the hypothesis that the mean life of bulbs exceed 850 hours. Use $\alpha = 0.05$.
(ii) Construct a 99% confidence interval on the mean life of bulbs.

**1.5** The output voltage measured from two brands of compressors *A* and *B* is as follows. The samples were selected randomly.

Brand A: 230, 225, 220, 250, 225, 220, 220, 230, 240, 245
Brand B: 220, 215, 222, 230, 240, 245, 230, 225, 250, 240

Assume that the output voltage follows normal distribution and has equal variances.
(i) Test the hypothesis that the output voltage from both the brands is same. Use $\alpha = 0.05$.
(ii) Construct a 95% confidence interval on the difference in the mean output voltage.

**1.6** The percentage shrinkage of two different types of casings selected randomly is given below:

Type 1: 0.20, 0.28, 0.12, 0.20, 0.24, 0.22, 0.32, 0.26, 0.32, 0.20
Type 2: 0.10, 0.12, 0.15, 0.09, 0.20, 0.12, 0.14, 0.16, 0.18, 0.20

Assume that the shrinkage follows normal distribution and has unequal variances.

(i) Test the hypothesis that Type 1 casing produces more shrinkage than Type 2 casing.
(ii) Find 90% confidence interval on the mean difference in shrinkage.

**1.7** The life of two brands of electric bulbs of same wattage measured in hours is given below:

Brand *X*: 800, 850, 900, 750, 950, 700, 800, 750, 900, 800

Brand *Y*: 950, 750, 850, 900, 700, 790, 800, 790, 900, 950

(i) Test the hypothesis that the lives of both the brands is same. Use $\alpha = 0.05$.
(ii) Construct 95% confidence interval on mean difference in lives of the bulbs.

**1.8** A training manager wanted to investigate the effect of a specific training programme on a group of participants. Before giving the training, he selected a random sample of 10 participants and administered a test. After the training programme he administered the same type of test on the same participants. The test scores obtained before and after the training are given below:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Before training: | 69 | 67 | 55 | 43 | 77 | 46 | 75 | 52 | 43 | 65 |
| After training: | 56 | 70 | 62 | 72 | 71 | 50 | 68 | 72 | 68 | 68 |

Test whether the training programme has any effect on the participants. Use $\alpha = 0.05$.

**1.9** A gym claims that their 15 week weight reduction exercise will significantly reduce weight. The table below gives the weight of 7 male adults selected randomly from a group undergoing the exercise.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Before: | 81 | 75 | 89 | 91 | 110 | 70 | 90 |
| After: | 75 | 73 | 87 | 85 | 90 | 65 | 80 |

Using the 5% significance level, determine whether the weight reduction exercise has any effect.

CHAPTER 2

# Fundamentals of Experimental Design

## 2.1 INTRODUCTION

Experimentation is one of the most common activities performed by everyone including scientists and engineers. It covers a wide range of applications from agriculture, biological, social science, manufacturing and service sectors, etc. Experimentation is used to understand and/or improve a system. The system may be a simple/complex product or process. A product can be developed in engineering, biology or physical sciences. A process can be a manufacturing process or service process such as health care, insurance, banking, etc. Experimentation can be used for developing new products/processes as well as for improving the quality of existing products/processes. In any experimentation the investigator tries to find out the effect of input variables on the output /performance of the product/process. This enables the investigator to determine the optimum settings for the input variables.

## 2.2 EXPERIMENTATION

Experimental design is a body of knowledge and techniques that assist the experimenter to conduct experiment economically, analyse the data, and make connections between the conclusions from the analysis and the original objectives of the investigation. Although the major emphasis in this book is on engineering products/processes, the methods can be applied to all other disciplines also.

The traditional approach in the industrial and the scientific investigation is to employ trial and error methods to verify and validate the theories that may be advanced to explain some observed phenomenon. This may lead to prolonged experimentation and without good results. Some of the approaches also include one factor at a time experimentation, several factors one at a time and several factors all at the same time. These approaches are explained in the following sections.

### 2.2.1 Conventional Test Strategies

***One-factor experiments:*** The most common test plan is to evaluate the effect of one parameter on product performance. In this approach, a test is run at two different conditions of that parameter (Table 2.1). Suppose we have several factors including factor $A$. In this strategy we keep all other factors constant, and determine the effect of one factor (say $A$). The first trial is conducted with factor $A$ at first level and the second trial is with level two. If there is any change in the result (difference in the average between $Y_1$ and $Y_2$), we attribute that to the factor $A$. If the first factor chosen fails to produce the expected result, some other factors would be tested.

**TABLE 2.1** One factor at a time experiment

| *Trial* | *Factor level* | *Test result* | *Test average* |
|---|---|---|---|
| 1 | $A_1$ | * * | $Y_1$ |
| 2 | $A_2$ | * * | $Y_2$ |

***Several factors one at a time:*** Suppose we have four factors ($A$, $B$, $C$, $D$). In this strategy (Table 2.2), the first trial is conducted with all factors at their first level. This is considered to be *base level experiment*. In the second trial, the level of one factor is changed to its second level (factor $A$). Its result ($Y_2$) is compared with the base level experiment ($Y_1$). If there is any change in the result (between $Y_1$ and $Y_2$), it is attributed to factor $A$. In the third trial, the level of factor $B$ is changed and the test average $Y_3$ is compared with $Y_1$ to determine the effect of $B$. Thus, each factor level is changed one at a time, keeping all the other factors constant. This is the traditional scientific approach to experimentation. In this strategy, the combined effect due to any two/more factors cannot be determined.

**TABLE 2.2** Several factors one at a time strategy

| *Trial* | *Factors* | | | | *Test result* | *Test average* |
|---|---|---|---|---|---|---|
| | *A* | *B* | *C* | *D* | | |
| 1 | 1 | 1 | 1 | 1 | * * | $Y_1$ |
| 2 | 2 | 1 | 1 | 1 | * * | $Y_2$ |
| 3 | 1 | 2 | 1 | 1 | * * | $Y_3$ |
| 4 | 1 | 1 | 2 | 1 | * * | $Y_4$ |
| 5 | 1 | 1 | 1 | 2 | * * | $Y_5$ |

***Several factors all at the same time:*** The third and most urgent situation finds the experimenter changing several things all at the same time with a hope that at least one of these changes will improve the situation sufficiently (Table 2.3). In this strategy, the first trial is conducted with all factors at first level. The second trial is conducted with all factors at their second level. If there

**TABLE 2.3** Several factors all at the same time strategy

| *Trial* | *Factors* | | | | *Test result* | *Test average* |
|---|---|---|---|---|---|---|
| | *A* | *B* | *C* | *D* | | |
| 1 | 1 | 1 | 1 | 1 | * * * | $Y_1$ |
| 2 | 2 | 2 | 2 | 2 | * * * | $Y_2$ |

is any change in the test average between $Y_1$ and $Y_2$, it is not possible to attribute this change in result to any of the factor/s. It is not at all a useful test strategy.

These are poor experimental strategies and there is no scientific basis. The effect of interaction between factors cannot be studied. The results cannot be validated. Hence, there is a need for scientifically designed experiments.

### 2.2.2 Better Test Strategies

A full factorial design with two factors $A$ and $B$ each with two levels will appear as given in Table 2.4.

**TABLE 2.4** Full factorial experiment

| *Trial* | *Factors and factor levels* | | |
|---|---|---|---|
| | *A* | *B* | *Response* |
| 1 | 1 | 1 | * * |
| 2 | 1 | 2 | * * |
| 3 | 2 | 1 | * * |
| 4 | 2 | 2 | * * |

Here, one can see that the full factorial experiment is orthogonal. There is an equal number of test data points under each level of each factor. Because of this balanced arrangement, factor $A$ does not influence the estimate of the effect of factor $B$ and vice versa. Using this information, both factor and interaction effects can be estimated.

A full factorial experiment is acceptable when only a few factors are to be investigated. When several factors are to be investigated, the number of experiments to be run under full factorial design is very large.

### 2.2.3 Efficient Test Strategies

Statisticians have developed more efficient test plans, which are referred to as fractional factorial experiments. These designs use only a portion of the total possible combinations to estimate the main factor effects and some, not all, of the interactions. The treatment conditions are chosen to maintain the orthogonality among the various factors and interactions.

Taguchi developed a family of fractional factorial experimental matrices, called Orthogonal Arrays (OAs) which can be utilized under various situations. One such design is an $L_8$ OA, with only 8 of the possible 128 treatment combinations. This is actually a one-sixteenth fractional factorial design.

## 2.3 NEED FOR STATISTICALLY DESIGNED EXPERIMENTS

The main reason for designing the experiment statistically is to obtain unambiguous results at a minimum cost. The statistically designed experiment permits simultaneous consideration of all possible variables/factors that are suspected to have a bearing on the problem under consideration and as such even if interaction effects exist, a valid evaluation of the main factors can be made. From a limited number of experiments the experimenter would be able to uncover the vital factors on which further trials would lead him to track down the most desirable combination of factors which will yield the expected results. The statistical principles employed in the design and analysis of experimental results assure impartial evaluation of the effects on an objective basis. The statistical concepts used in the design form the basis for statistically validating the results from the experiments.

## 2.4 ANALYSIS OF VARIANCE

The statistical foundations for design of experiments and the Analysis of Variance (ANOVA) was first introduced by Sir Ronald A, Fisher, the British biologist. ANOVA is a method of partitioning total variation into accountable sources of variation in an experiment. It is a statistical method used to interpret experimented data and make decisions about the parameters under study.

The basic equation of ANOVA is given by

Total sum of squares = sum of squares due to factors + sum of squares due to error

$$SS_{\text{Total}} = SS_{\text{Factors}} + SS_{\text{Error}}$$

### *Degrees of freedom*

A degree of freedom in a statistical sense is associated with each piece of information that is estimated from the data. For instance, mean is estimated from all data and requires one degree of freedom (df) for that purpose. Another way to think the concept of degree of freedom is to allow 1df for each fair (independent) comparison that can be made in the data. Similar to the total sum of squares, summation can be made for degrees of freedom.

Total df = df associated with factors + error df

### *Variance*

The variance ($V$) of any factor/component is given by its sum of squares divided by its degrees of freedom. It is also referred to as Mean Square (MS).

### *F-test*

$F$-test is used to test whether the effects due to the factors are significantly different or not. $F$-value is computed by dividing the factor variance/mean square with error variance/mean square. However, in the case of multi-factor designs with random effects or mixed effects model, the denominator for computing $F$-value shall be determined by computing expected mean squares.

### *ANOVA table*

A summary of all analysis of variance computations are given in the ANOVA table. A typical format used for one factor is given in Table 2.5.

**TABLE 2.5** Analysis of variance computations (ANOVA)

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean square/ variance* | $F_0$ |
|---|---|---|---|---|
| Factor | $SS_F$ | $K - 1$ | $V_F = SS_F/K - 1$ | $V_F/V_e$ |
| Error | $SS_e$ | $N - K$ | $V_e = SS_e/N - K$ | |
| Total | $SS_{\text{Total}}$ | $N - 1$ | | |

where

$N$ = total number of observations
$SS_F$ = sum of squares of factor
$K$ = number of levels of the factor
$SS_e$ = sum of squares of error
$F_0$ = computed value of $F$
$V_F$ = variance of factor
$V_e$ = variance of error

## 2.5 BASIC PRINCIPLES OF DESIGN

The purpose of statistically designed experiments is to collect appropriate data which shall be analysed by statistical methods resulting in valid and objective conclusions. The two aspects of experimental problem are as follows:

1. The design of the experiment, and
2. The statistical analysis of the data.

The following basic principles are used in planning/designing an experiment.

### 2.5.1 Replication

Replication involves the repetition of the experiment and obtaining the response from the same experimental set up once again on different experimental unit (samples). An experimental unit may be a material, animal, person, machine, etc. The purpose of replication is to obtain the magnitude of experimental error. This error estimate (error variance) is used for testing statistically the observed difference in the experimental data. Replications also permit the experimenter to obtain a precise estimate of the effect of a factor studied in the experiment. Finally, it is to be noted that replication is not a repeated measurement. Suppose in an experiment five hardness measurements are obtained on five samples of a particular material using the same tip for making the indent. These five measurements are five replicates.

### 2.5.2 Randomization

The use of statistical methods requires randomization in any experiment. The allocation of experimental units (samples) for conducting the experiment as well as the order of experimentation should be random. Statistical methods require that the observations (or errors) are independently distributed random variables. It meets this requirement. It also assists in averaging out the effects of extraneous factors that may be present during experimentation. When complete randomization is not possible, appropriate statistical design methods shall be used to tackle restriction on randomization.

### 2.5.3 Blocking

Blocking is a design technique used to improve the precision of the experiment. Blocking is used to reduce or eliminate the effect of nuisance factors or noise factors. A block is a portion of the experimental material that should be more homogeneous than the entire set of material or a block is a set of more or less homogeneous experimental conditions. It is also a restriction on complete randomization. More about blocking will be discussed in factorial designs.

The three basic principles of experimental design, randomization, replication and blocking are part of all experiments.

## 2.6 TERMINOLOGY USED IN DESIGN OF EXPERIMENTS

***Factor:*** A variable or attribute which influences or is suspected of influencing the characteristic being investigated. All input variables which affect the output of a system are factors. Factors are varied in the experiment. They can be controlled at fixed levels. They can be varied or set at levels of our interest. They can be qualitative (type of material, type of tool, etc.) or quantitative (temperature, pressure, etc.). These are also called *independent variables*.

***Levels of a factor:*** The values of a factor/independent variable being examined in an experiment. If the factor is an attribute, each of its state is a level. For example, setting of a switch on or off are the two levels of the factor switch setting. If the factor is a variable, the range is divided into required number of levels. For example, the factor temperature ranges from 1000 to 1200°C and it is to be studied at three values say 1000°C, 1100°C and 1200°C, these three values are the three levels of the factor temperature. The levels can be fixed or random.

***Treatment:*** One set of levels of all factors employed in a given experimental trial. For example, an experiment conducted using temperature $T_1$ and pressure $P_1$ would constitute one treatment. In the case of single factor experiment, each level of the factor is a treatment.

***Experimental unit:*** Facility with which an experimental trial is conducted such as samples of material, person, animal, plant, etc.

***Response:*** The result/output obtained from a trial of an experiment. This is also called dependent variable. Examples are yield, tensile strength, surface finish, number of defectives, etc.

***Effect:*** Effect of a factor is the change in response due to change in the level of the factor.

***Experimental error:*** It is the variation in response when the same experiment is repeated, caused by conditions not controlled in the experiment. It is estimated as the residual variation after the effects have been removed.

## 2.7 STEPS IN EXPERIMENTATION

The experimenter must clearly define the purpose and scope of the experiment. The objective of the experiment may be, for example to determine the optimum operating conditions for a process to evaluate the relative effects on product performance of different sources of variability or to determine whether a new process is superior to the existing one. The experiment will be planned properly to collect data in order to apply the statistical methods to obtain valid and objective conclusions. The following seven steps procedure may be followed:

1. Problem statement
2. Selection of factors, levels and ranges
3. Selection of response variable
4. Choice of experimental design
5. Conducting the experiment
6. Analysis of data
7. Conclusions and recommendations

### 2.7.1 Problem Statement

Initially, whenever there is problem of inferior process or product performance, all the traditional quality tools will be tried. Still if the problem persists, to improve the performance further, design of experiments would be handy. Since experimentation takes longer time and may disturb the regular production, problem selection should be given importance. Problem should be clearly defined along with its history if any. A clear definition of the problem often contributes to better understanding of the phenomenon being studied and the final solution of the problem. The various questions to be answered by the experimenter and the expected objectives may also be formulated. The insight into the problem leads to a better design of the experiment. When we state the problem, we have to be very specific. A few examples of specifying a problem are given below:

(i) If there is a rejection in bore diameter, we need to specify whether the rejection is due to oversize or undersize or both.
(ii) Suppose there is a variation in the compression force of a shock absorber. We have to specify clearly whether the problem is less force or more force or both.
(iii) If the problem is a defect like crack, blister, etc., we should also specify whether the problem is observed as random phenomenon on the product or concentrated to one specific area.

### 2.7.2 Selection of Factors, Levels and Ranges

There may be several factors which may influence the performance of a process or product. These factors include the design factors which can be controlled during experimentation. Identification of these factors leads to the selection of experimental design. Here, the customer expectations should also be kept in mind. The following methods may be used:

(i) Brainstorming
(ii) Flowcharting (especially for processes)
(iii) Cause-effect diagrams

Often, the experience and the specialized knowledge of engineers and scientists dominate the selection of factors.

Initial rounds of experimentation should involve many factors at few levels; two are recommended to minimize the size of the beginning experiment. Initial experiments will eliminate many factors from contention and the few remaining can then be investigated with multiple levels.

Depending on the availability of resources and time for experimentation, the number of levels for each factor is usually decided. In order to estimate factor effect, a minimum of two levels should be used. In preliminary experimentation, where large number of factors is included in the study, the number of levels is limited to two. In detailed experimentation with a few factors, more number of levels is considered. The range over which each factor to be studied and the values for each level are fixed taking into account the working range of the process. Usually, this requires the process knowledge and past experience.

Selection of levels for qualitative variables is not difficult. Only thing is that the level must be practical and should give valid data. Sometimes 'on' and 'off' or 'yes' and 'no' can also be the two qualitative levels of a factor. The choice of levels for quantitative factors is often critical. Over the operating range of the variable which of the values to be considered as the levels has to be carefully judged. Selection of the extreme (within the operating range) is always safe because the system would function safely. In doing so one may miss the range that might give best response. However, if levels are chosen too close together, one may not see any difference and may miss something important outside the range of experimentation. So it is always better to include experts and operators who are knowledgeable in the process to suggest levels for study. And it is preferable to include the current operating value as one of the levels in experimentation.

### 2.7.3 Selection of Response Variable

The response variable selected must be appropriate to the problem under study and provide useful information. A clear statement is required of the response to be studied, i.e., the dependent variables to be evaluated. Frequently there may be a number of such response variables, for example, tensile strength, surface finish, % elongation, leakages, % defectives, etc. It is important that standard procedure for measuring such variables should be established and documented. For visual and sensory characteristics, it is particularly important that standards be developed to remove subjectivity in judgments made by different observers of different times. The capability of the measuring devices must also be ensured.

### 2.7.4 Choice of Experimental Design

Several experimental designs are available like factorial designs (single factor and multifactor), fractional factorial designs, confounding designs, etc. Selecting the right type of design for the problem under study is very important. Otherwise the results will be misleading. Choice of design involves the consideration of the number of factors and their levels, resources required such as sample size (number of replications), availability of experimental units (samples), time to complete one replication, randomization restrictions applicability of blocking, etc. In selecting the design it is important to keep the objectives in mind. In this book some important designs widely applied are discussed.

### 2.7.5 Conducting the Experiment

It is to be noted that the test sheets should show only the main factor levels required for each trial. Only the analysis is concerned with the interaction columns. When conducting the experiment, it is important to monitor the process to ensure that everything is being carried out as per the plan. The allocation of experimental units as well as the order of experimentation of the experimental trials should be random. This will even out the effect of unknown and controlled factors that may vary during the experimentation period.

Randomization may be complete, simple repetition or complete within blocks.

*Complete randomization* means any trial has an equal chance of being selected for the first test. Using random numbers, a trial can be selected. This method is used when test set up change is easy and inexpensive.

*Simple repetition* means that any trial has an equal chance of being selected for the first test, but once that trial is selected all the repetitions are tested for that trial. This method is used if test set ups are very difficult or expensive to change.

*Complete randomization within blocks* is used, where one factor may be very difficult or expensive to change the test set up for, but others are very easy. If factor $A$ were difficult to change, the experiment could be completed in two blocks. All $A_1$ trials could be selected randomly and then all $A_2$ trials could be randomly selected.

The different methods of randomization affect error variance in different ways. In simple repetition, trial to trial variation is large and repetition variation is less. This may cause certain factors in ANOVA to be significant when in fact they are not. Hence, complete randomization is recommended.

### 2.7.6 Analysis of Data

Several methods are available for analysing the data collected from experiment. However, statistical methods should be used to analyse the data so that results and conclusions are objective. Analysis of variance is widely used to test the statistical significance of the effects through $F$-Test. Confidence interval estimation is also part of the data analysis. Often an empirical model is developed relating the dependent (response) and independent variables (factors). Residual analysis and model adequacy checking are also part of the data analysis procedure. Statistical analysis of data is a must for academic and scientific purpose. In industrial experiments, the graphical analysis and normal probability plot of the effects may be preferred.

### 2.7.7 Conclusions and Recommendations

After the data are analysed using the statistical methods, the experimenter must interpret the results from the practical point of view and recommend the same for possible implementation. The recommendations include the settings or levels for all the factors (input variables studied) that optimizes the output (response). It is always better to conduct conformation tests using the recommend levels for the factors to validate the conclusions.

## 2.8 CHOICE OF SAMPLE SIZE

### 2.8.1 Variable Data

The selection of sample size (number of repetitions) is important. A minimum of one test result for each trial is required. More than one test per trial increases the sensitivity of experiment to detect small changes in averages of populations. The sample size (the number of replications) depends on the resources available (the experimental units) and the time required for conduct of the experiment. The sample size should be as large as possible so that it meets the statistical principles and also facilitate a better estimate of the effect. Often the number of replicates is chosen arbitrarily based on historical choices or guidelines.

### 2.8.2 Attribute Data

Attribute data provides less discrimination than variable data. That is, when an item is classified as bad (defective), the measure of how bad is not indicated. Because of this reduced discrimination, many pieces of attribute data are required. A general guideline is that the class (occurrence or non-occurrence) with the least frequency should have at least a count of 20. For example, in a study on defective parts, we should obtain at least a total of 20 defectives. Suppose the past per cent defective in a process is 5%. Then the total sample size for the whole study (all trials/ runs) shall be 400 to obtain an expected number of 20 defectives.

## 2.9 NORMAL PROBABILITY PLOT

One of the assumptions in all the statistical models used in design of experiments is that the experimental errors (residuals) are normally distributed. This normality assumption can be verified by plotting the residuals on a normal probability paper. If all the residuals fall along a straight line drawn through the plotted points, it is inferred that the residuals/errors follow normal distribution. Also normal probability plot of effects is used to judge the significance of effects. The effects which fall away from the straight line are judged as significant effects. This method is often used by practitioners in the industry.

Alternatively we can use half-normal probability plot of residuals and effects. The absolute value of normal variable is half-normal. An advantage of half-normal plot is that all the large estimated effects appear in the right corner and fall away from the line. Sometimes normal probability plots misleads in identifying the real significant effects which is avoided in half-normal probability plots.

### 2.9.1 Normal Probability Plotting

The sample data is arranged in ascending order of the value of the variable ($X_1, X_2, \ldots, X_j, \ldots, X_n$). And the ordered data ($X_j$) is plotted against their observed cumulative frequency $(j - 0.5)/n$ on the normal probability paper. If all the data points fall approximately along a straight line, it is concluded that the data (variable) follows normal distribution. The mean is estimated as the

50th percentile value on the probability plot. And the standard deviation is given by the difference between 84th percentile and the 50th percentile value.

If residuals are plotted on the normal probability paper, the data corresponding to those residuals which fall away from the straight line are called outliers and such data shall not be considered. If the factor effects are plotted on the normal probability paper, the effects which fall away from the straight line are concluded as significant effects.

### 2.9.2 Normal Probability Plot on an Ordinary Graph Paper

Arrange the sample data in ascending order of the value of the variable ($X_1, X_2, \ldots, X_n$).

Compute the cumulative frequency

$$\frac{j - 0.5}{n} \tag{2.1}$$

Transform the cumulative frequency into a standardized normal score $Z_j$

$$Z_j = \phi\left(\frac{j - 0.5}{n}\right) \qquad j = 1, 2, 3, \ldots, n \tag{2.2}$$

Plot $Z_j$ vs $X_j$

The interpretation of this plot is same as explained in Section 2.9.1 above.

***Normal probability plot of data***

The following data represent the hardness of 10 samples of certain alloy steel measured after a heat treatment process. How to obtain a normal probability plot on a normal probability paper is illustrated below:

275, 258, 235, 228, 265, 223, 261, 200, 276, 237

Table 2.6 shows the computations required for normal probability plot. The data arranged in the ascending order of value is shown in the second column of Table 2.6. The first column indicates the rank of the ordered data. Third column gives the cumulative frequency of the ordered data. The last column ($Z_j$) is the standardized normal score.

**TABLE 2.6** Computations for normal probability plot

| $j$ | $X_j$ | $(j - 0.5)/10$ | $Z_j$ |
|---|---|---|---|
| 1 | 195 | 0.05 | –1.64 |
| 2 | 223 | 0.15 | –1.04 |
| 3 | 228 | 0.25 | –0.67 |
| 4 | 235 | 0.35 | –0.39 |
| 5 | 237 | 0.45 | –0.13 |
| 6 | 248 | 0.55 | 0.13 |
| 7 | 256 | 0.65 | 0.39 |
| 8 | 263 | 0.75 | 0.67 |
| 9 | 272 | 0.85 | 1.04 |
| 10 | 290 | 0.95 | 1.64 |

The plot $X_j$ ($X$-axis) versus $(j - 0.5)/10$ ($Y$-axis) on a normal probability paper is the normal probability plot of the data (Figure 2.1). We can also plot $X_j$ ($X$-axis) versus $Z_j$ ($Y$-axis) on an ordinary graph sheet to obtain the same result. This is left as an exercise to the reader.

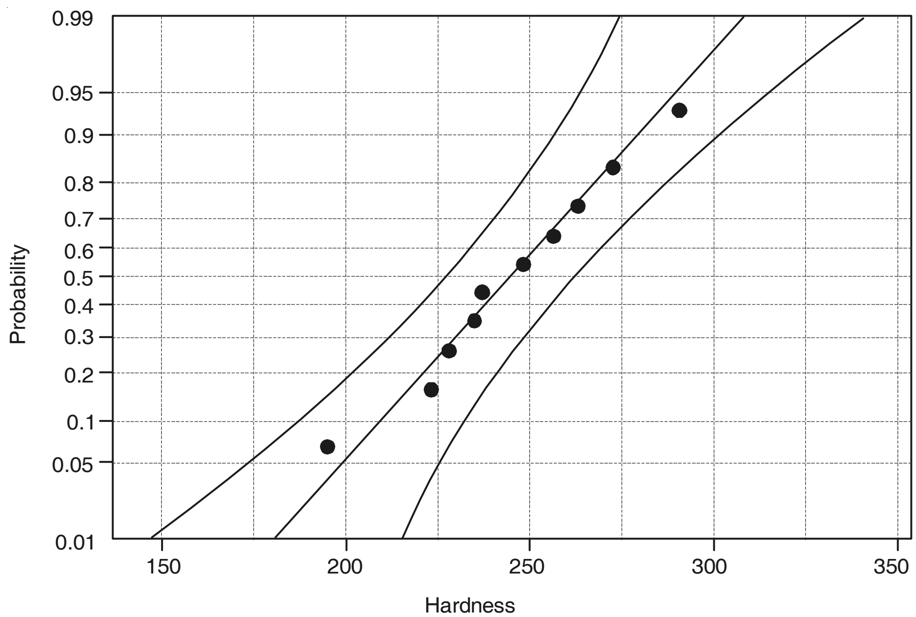



**FIGURE 2.1** Normal probability plot of data.

#### *Normal probability plot of effects*

Suppose we have the following factor effects from an experiment.

| | | |
|---|---|---|
| $a = -0.2875$ | $ab = 0.4375$ | $abc = 0.5875$ |
| $b = 0.7625$ | $ac = 0.1625$ | $abd = -0.4125$ |
| $c = 0.2375$ | $ad = 0.5625$ | $acd = 0.2125$ |
| $d = -1.3625$ | $bd = -0.5875$ | $bcd = -0.3375$ |
| $bc = -0.7875$ | $cd = -0.6125$ | $abcd = 0.5375$ |

The computations required for normal probability plot of the above effects are given in Table 2.7. The cumulative normal probability is plotted on a normal probability paper against the effects. Figure 2.2 shows the normal probability plot of effects. The effects corresponding to the points which fall away from the straight line are considered to be *significant*. From Figure 2.2 we observe that two points fall away from the straight line drawn through the points.

### 2.9.3 Half-normal Probability Plotting

The half-normal probability plot is generally used for plotting the factor effects. This is a plot of the absolute values of the effects against their cumulative normal probability. This plot is particularly

**TABLE 2.7** Computations for normal probability plot of effects

| *j* | *Effects* ($X_j$) | *Normal probability* $(j - 0.5)/15$ |
|---|---|---|
| 1 | $d = -1.3625$ | 0.0333 |
| 2 | $bc = -0.7875$ | 0.1000 |
| 3 | $cd = -0.6125$ | 0.1666 |
| 4 | $bd = -0.5875$ | 0.2333 |
| 5 | $abd = -0.4125$ | 0.3000 |
| 6 | $bcd = -0.3375$ | 0.3666 |
| 7 | $a = -0.2875$ | 0.4333 |
| 8 | $ac = 0.1625$ | 0.5000 |
| 9 | $acd = 0.2125$ | 0.5666 |
| 10 | $c = 0.2375$ | 0.6333 |
| 11 | $ab = 0.4375$ | 0.7000 |
| 12 | $abcd = 0.5375$ | 0.7666 |
| 13 | $ad = 0.5625$ | 0.8333 |
| 14 | $abc = 0.5875$ | 0.9000 |
| 15 | $b = 0.7625$ | 0.9666 |

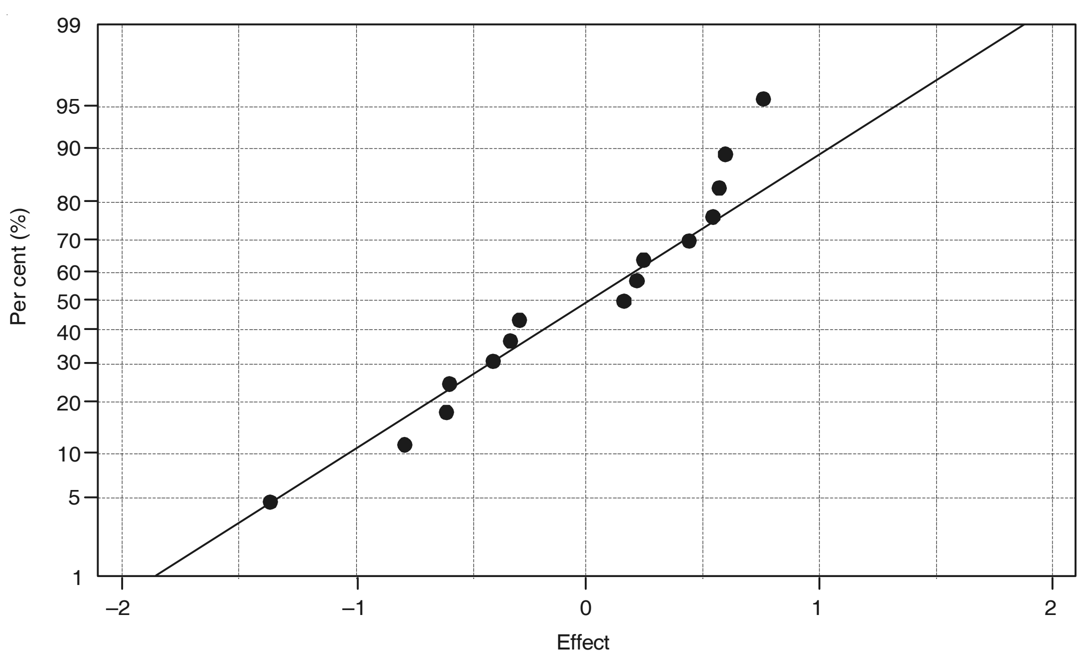



**FIGURE 2.2** Normal plot of the effects.

useful when there are only few effects such as an eight run design. We consider the absolute effects and arrange them in the increasing order and the observed cumulative frequency $(j - 0.5)/n$ is plotted on a half-normal probability paper.

Alternately the ordered effects (Table 2.7) are plotted against their observed cumulative

$$\text{Probability} = 0.5 + 0.5\left(\frac{j - 0.5}{n}\right) \qquad j = 1, 2, 3, \ldots, n \tag{2.3}$$

on a normal probability paper.

This probability can be converted into a standardized normal score $Z_j$ and plot on an ordinary graph paper,

where
$$Z_j = \phi\left[0.5 + 0.5\left(\frac{j - 0.5}{n}\right)\right] \qquad j = 1, 2, 3, \ldots, n \tag{2.4}$$

This is left as an exercise to the reader.

The advantage of half-normal plot is that all the large estimated effects appear in the right corner and fall away from the line. The effects which fall away from the straight line are considered to be significant effects.

### *Half-normal probability plot of effects*

Consider the same effects used for normal plot (Table 2.7). Here we consider the absolute effects and arrange them in increasing order and the normal probability is computed. This probability is plotted on a half-normal probability paper. The required computations are given in Table 2.8. Figure 2.3 shows the half-normal plot of the effects. From this half-normal plot it is seen that

**TABLE 2.8** Computations for half-normal probability plot

| $j$ | *Effects* $\lvert X_j \rvert$ | *Normal probability* $(j - 0.5)/16$ |
|---|---|---|
| 1 | 0.1625 | 0.0333 |
| 2 | 0.2125 | 0.1000 |
| 3 | 0.2375 | 0.1666 |
| 4 | 0.2875 | 0.2333 |
| 5 | 0.3375 | 0.3000 |
| 6 | 0.4125 | 0.3666 |
| 7 | 0.4375 | 0.4333 |
| 8 | 0.5375 | 0.5000 |
| 9 | 0.5625 | 0.5666 |
| 10 | 0.5875 | 0.6333 |
| 11 | 0.5875 | 0.7000 |
| 12 | 0.6125 | 0.7666 |
| 13 | 0.7625 | 0.8333 |
| 14 | 0.7875 | 0.9000 |
| 15 | 1.3625 | 0.9666 |

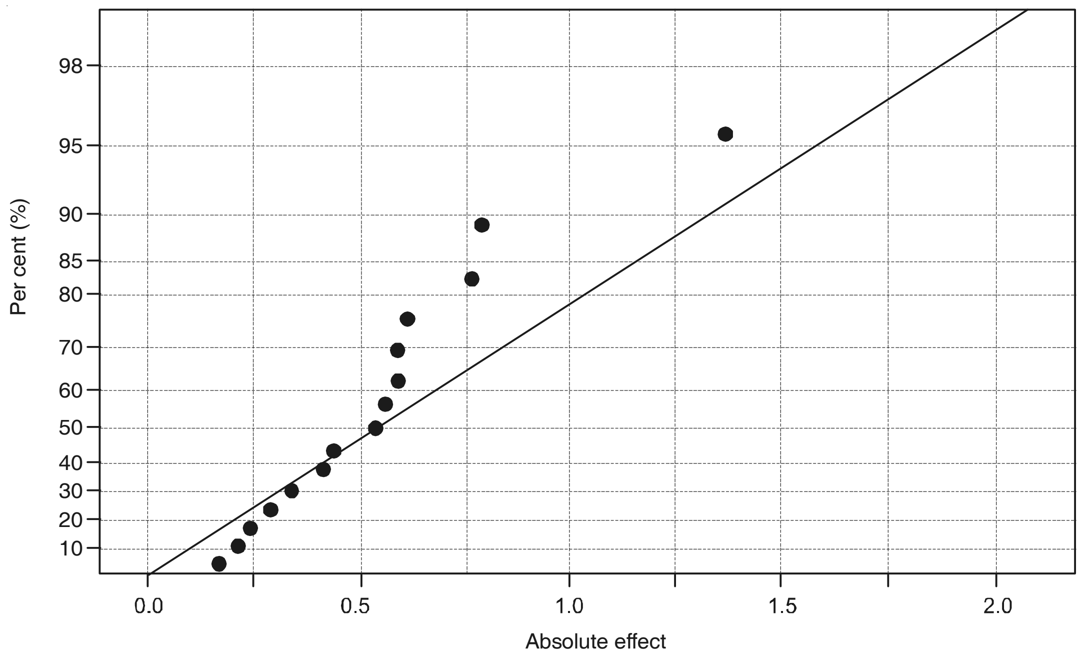



**FIGURE 2.3** Half-normal plot of the effects.

only one point (effect) lies far away from the straight line which is considered to be *significant*. Thus, there is a marked difference between normal and half-normal plot. It is recommended to use half-normal plot to identify significant effects.

## 2.10 BRAINSTORMING

Brainstorming is an activity which promotes group participation and team work, encourages creative thinking and stimulates the generation of as many ideas as possible in a short period of the time. The participants in a brainstorming session are those who have the required domain knowledge and experience in the area associated with the problem under study. An atmosphere is created such that everybody feels free to express themselves.

All ideas from each participant are recorded and made visible to all the participants. Each input and contribution is recognized as important and output of the whole session is in the context.

The continuing involvement of each participant is assured and the groups future is reinforced by mapping out the exact following actions (analysis and evaluation of ideas) and the future progress of the project.

## 2.11 CAUSE AND EFFECT ANALYSIS

Cause and effect analysis is a technique for identifying the most probable causes (potential cause) affecting a problem. The tool used is called cause and effect diagram. It can help to analyse cause and effect relationships and used in the conjunction with brainstorming.

Cause and effect diagram visually depicts a clear picture of the possible causes of a particular effect. The effect is placed in a box on the right and a long process line is drawn pointing to the box (Figure 2.4). The effect may be process rejections or inferior performance of a product, etc. After deciding the major categories of causes, these are recorded on either side of the line within boxes connected with the main process line through other lines. Usually, the major categories of causes recorded are due to person (operator), machine (equipment), method (procedure) and material. Each one of these major causes can be viewed as an effect in its own right with its own process line, around which other associated causes are clustered.

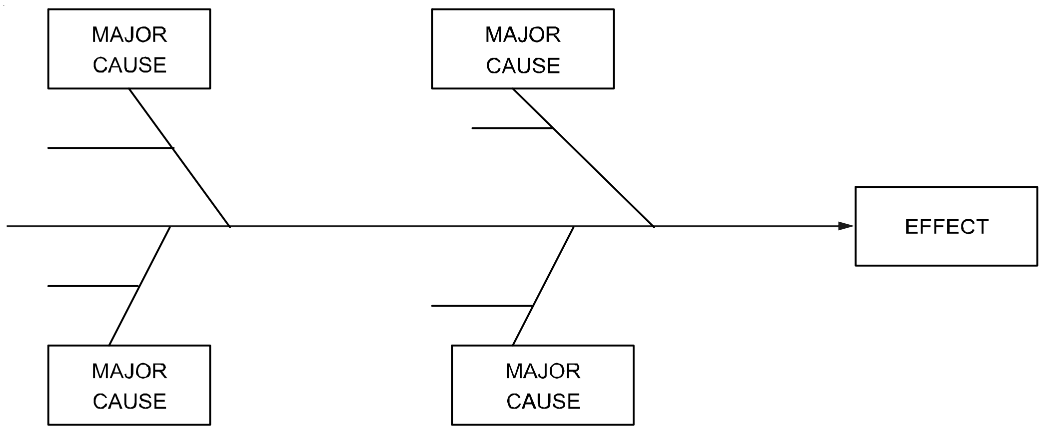



**FIGURE 2.4** Cause and effect diagram.

These minor causes are identified through brainstorming. The experimenter, through discussion and process of elimination arrive at the most likely causes (factors) to be investigated in experimentation. The cause and effect diagram is also known as *Fish Bone diagram* or *Ishikawa diagram* (named after Japanese professor Ishikawa).

## 2.12 LINEAR REGRESSION

When experiments are conducted involving only quantitative factors like temperature, pressure, concentration of a chemical, etc., often the experimenter would like to relate the output (response) with the input variables (factors) in order to predict the output or optimize the process. If the study involves only one dependent variable (response), one independent variable (factor) and if the relation between them is linear, it is called *simple linear regression*. If the response is related linearly with more than one independent variable, it is called *multiple linear regression*.

### 2.12.1 Simple Linear Regression Model

Suppose the yield ($Y$) of a chemical process depends on the temperature ($X$). If the relation between $Y$ and $X$ is linear, the model that describes this relationship is

$$Y = \beta_0 + \beta_1 X + e \qquad (2.5)$$

where $\beta_0$ and $\beta_1$ are the constants called regression coefficients and $e$ is a random error, normally independently distributed with mean zero and variance $\sigma_e^2$, NID (0, $\sigma^2$)

$\beta_0$ is also called the intercept and $\beta_1$ is the slope of the line, that is fitted to the data. If we have $n$ pairs of data ($X_1$, $Y_1$), ($X_2$, $Y_2$), … ($X_n$, $Y_n$), we can estimate the parameters by the method of least squares. The model with the sample data Eq. (2.6) is of the same form as above Eq. (2.5).

$$Y_i = b_0 + b_1 X_i + e_i \tag{2.6}$$

where $b_0$ and $b_1$ are the estimates of $\beta_0$ and $\beta_1$ respectively. Applying the method of least squares we get the following two normal equations which can be solved for the parameters.

$$\sum_{i=1}^{n} Y_i = nb_0 + b_1 \sum_{i=1}^{n} X_i \tag{2.7}$$

$$\sum_{i=1}^{n} X_i Y_i = b_0 \sum_{i=1}^{n} X_i + b_1 \sum_{i=1}^{n} X_i^2 \tag{2.8}$$

$$b_1 = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2} \tag{2.9}$$

$$b_0 = \bar{Y} - b_1 \bar{X} \tag{2.10}$$

where $\bar{X} = \dfrac{\sum X_i}{n}$ and $\bar{Y} = \dfrac{\sum Y_i}{n}$.

From the normal Eqs. (2.7) and (2.8), the parameters $b_0$ and $b_1$ can also be expressed in terms of sums of squares as follows:

$$b_1 = SS_{XY}/SS_X \tag{2.11}$$

$$b_0 = \bar{Y} - b_1 \bar{X} \tag{2.12}$$

$$SS_X = \sum (X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2 \tag{2.13}$$

$$SS_Y = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - n\bar{Y}^2 \tag{2.14}$$

$$SS_{XY} = \sum (X_i - \bar{X})(Y_i - \bar{Y}) = \sum X_i Y_i - (\sum X_i)(\sum Y_i)/n = \sum X_i Y_i - n\bar{X}\bar{Y} \tag{2.15}$$

It is to be noted that $SS_X$ and $SS_Y$ are the terms used to determine the variance of $X$ and variance of $Y$ respectively. $SS_X$ and $SS_Y$ are called corrected sum of squares. And $\beta_1$ and $\beta_0$ are estimated as follows:

$$\beta_1 = SS_{XY}/SS_X \tag{2.16}$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X} \text{ and} \tag{2.17}$$

$$Y = \beta_0 + \beta_1 X \text{ is the regression equation} \tag{2.18}$$

$$\text{Variance of } X(S_X^2) = SS_X/(n - 1) \text{ and} \tag{2.19}$$

$$\text{Variance of } Y(S_Y^2) = SS_Y/(n - 1) \tag{2.20}$$

Similarly, we can write error sum of squares ($SS_e$) for the regression as

$$SS_e = \sum e_i^2 \text{ and standard error as } S_e = \sqrt{\frac{\sum e_i^2}{n-2}} = \sqrt{\frac{SS_e}{\text{dfe}}} \tag{2.21}$$

where, dfe are error degrees of freedom.

The ANOVA equation for linear regression is

Total corrected sum of squares = sum of squares due to regression
+ sum of squares due to error

$$SS_{\text{Total}} = SS_R + SS_e \tag{2.22}$$

These sum of squares are computed as follows:

$$\text{Let CF (correction factor)} = \frac{(\text{Grand total})^2}{\text{Total number of observations}}$$

$$SS_{\text{Total}} = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} = \sum Y_i^2 - \text{CF} \tag{2.23}$$

$$SS_R = \beta_1 SS_{XY} \tag{2.24}$$

The ANOVA computations for simple linear regression are given in Table 2.9.

**TABLE 2.9** ANOVA for simple linear regression

| *Source* | *Sum of squares* | *Degree of freedom* | *Mean square* | *F* |
|---|---|---|---|---|
| Regression | $SS_R = \beta_1 SS_{XY}$ | 1 | $MS_R = SS_R/1$ | $MS_R/MS_e$ |
| Error | $SS_e = SS_{\text{Total}} - \beta_1 SS_{XY}$ | $n - 2$ | $MS_e = SS_e/n - 2$ | |
| Total | $SS_{\text{Total}}$ | $n - 1$ | | |

Test for significance of regression is to determine if there is a linear relationship between $X$ and $Y$.

The appropriate hypotheses are $H_0$: $\beta_1 = 0$

$H_1$: $\beta_1 \neq 0$

Reject $H_0$ if $F_0$ exceeds $F_{\alpha,1,n-2}$.

Rejection of $H_0$ implies that there is significant relationship between the variable $X$ and $Y$. we can also test the coefficients $\beta_0$ and $\beta_1$ using one sample $t$ test with $n - 2$ degrees of freedom.

The hypotheses are $H_0$: $\beta_0 = 0$

$H_1$: $\beta_0 \neq 0$

and $H_0$: $\beta_1 = 0$

$H_1$: $\beta_1 \neq 0$

The test statistics are

$$t_0 = \frac{\beta_0}{\sqrt{MS_e(1/n) + (\overline{X}^2/SS_X)}} \tag{2.25}$$

Reject $H_{0,}$ if $|t_0| > t_{\alpha/2,n-2}$

$$t_0 = \frac{\beta_1}{\sqrt{MS_e/SS_X}} \tag{2.26}$$

Reject $H_{0,}$ if $|t_0| > t_{\alpha/2,n-2}$

## ILLUSTRATION 2.1

**Simple Linear Regression**

A software company wants to find out whether their profit is related to the investment made in their research and development. They have collected the following data from their company records.

| | | | | | | |
|---|---|---|---|---|---|---|
| Investment in R&D (₹ in millions): | 2 | 3 | 5 | 4 | 11 | 5 |
| Annual profit (₹ in lakhs): | 20 | 25 | 34 | 30 | 40 | 31 |

(i) Develop a simple linear regression model to the data and estimate the profit when the investment is 7 million rupees.
(ii) Test the significance of regression using $F$-test.
(iii) Test significance of $\beta_1$.

**Solution:**

In this problem, the profit depends on investment on R&D.
Hence, $X$ = Investment in R&D and $Y$ = Profit
The summation values of various terms needed for regression are given in Table 2.10.

**TABLE 2.10** Computations for Illustration 2.1

| $X$ | $Y$ | $XY$ | $X^2$ | $Y^2$ |
|---|---|---|---|---|
| 2 | 20 | 40 | 4 | 400 |
| 3 | 25 | 75 | 9 | 625 |
| 5 | 34 | 170 | 25 | 1156 |
| 4 | 30 | 120 | 16 | 900 |
| 11 | 40 | 440 | 121 | 1600 |
| 5 | 31 | 155 | 25 | 961 |
| 30 | 180 | 1000 | 200 | 5642 |

$$\overline{X} = \frac{\sum X_i}{n} = \frac{30}{6} = 5.0$$

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{180}{6} = 30.0$$

$$SS_X = \Sigma(X_i - \overline{X})^2 = \Sigma X_i^2 - n\overline{X}^2 = 200 - 6(5)^2 = 50$$

$$SS_Y = \Sigma(Y_i - \overline{Y})^2 = \Sigma Y_i^2 - n\overline{Y}^2 = 5642 - 6(30)^2 = 242$$

$$SS_{XY} = \Sigma X_i Y_i - n\,\overline{X}\overline{Y} = 1000 - 6(5 \times 30) = 100$$

$$\hat{\beta}_1 = \frac{SS_{XY}}{SS_X} = \frac{100}{50} = 2.0$$

$$\hat{\beta}_0 = \overline{Y} - \beta_1 \overline{X} = 30 - (2 \times 5) = 20.0$$

The regression model $Y = \beta_0 + \beta_1 X$

$$= 20.0 + 2.0X$$

(i) When the investment is ₹ 7 million, the profit $Y = 20.0 + 2.0(7) = 34$ millions

(ii) $SS_{\text{Total}} = SS_Y = 242.0$

$SS_R = \beta_1 S_{XY} = 2.0(100) = 200.0$

$SS_e = SS_{\text{Total}} - SS_R = 242.0 - 200.0 = 42.0$

Table 2.11 gives ANOVA for Illustration 2.1.

**TABLE 2.11** ANOVA for Illustration 2.1 (Simple Linear Regression)

| *Source* | *Sum of squares* | *Degee of freedom* | *Mean square* | *F* |
|---|---|---|---|---|
| Regression | 200.0 | 1 | 200 | 19.05 |
| Error | 42.0 | 4 | 10.5 | |
| Total | 242.0 | 5 | | |

Since $F_{5\%,1,4} = 7.71$, regression is significant. That is the relation between $X$ and $Y$ is significant.

(iii) The statistic to test the regression coefficient $\beta_1$ is

$$t_0 = \frac{\beta_1}{\sqrt{\dfrac{MS_e}{SS_x}}} = \frac{2.0}{\sqrt{10.5/50}} = \frac{2}{0.458} = 4.367$$

Since $t_{0.025,4} = 2.776$, the regression coefficient $\beta_1$ is significant.

### 2.12.2 Multiple Linear Regression Model

In experiments if the response ($Y$) is linearly related to more than one independent variable, the relationship is modelled as multiple linear regression. Suppose, we have $X_1, X_2, X_3, \ldots, X_k$ independent variables (factors). A model that might describe the relationship is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + e \tag{2.27}$$

The parameters $\beta_j$, where, $j = 0, 1, 2, \ldots, k$ are called *regression coefficients*.

Models those are more complex in appearance than Eq. (2.27) may also be analysed by multiple linear regression. Suppose we want to add an interaction term to a first order model in two variables. Then the model is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + e \tag{2.28}$$

If we let $X_3 = X_1 X_2$ and $\beta_3 = \beta_{12}$, Eq. (2.28) can be written as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e \tag{2.29}$$

Similarly, a second order model in two variables is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 + e \tag{2.30}$$

can be modelled as a linear regression model Eq. (2.27)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + e \tag{2.31}$$

where $X_3 = X_1^2$, $X_4 = X_2^2$, $X_5 = X_1 X_2$
and $\beta_3 = \beta_{11}$, $\beta_4 = \beta_{22}$, $\beta_5 = \beta_{12}$

The parameters can be estimated by deriving least square normal equations and matrix approach.

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{2.32}$$

where, the $X$ matrix and $Y$ vector for $n$ pairs of data are

$$X = \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ 1 & X_{31} & X_{32} & \cdots & X_{3k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{pmatrix}_{n \times k+1} \tag{2.33}$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1} \tag{2.34}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_k \end{bmatrix}_{k+1\times 1} \tag{2.35}$$

The fitted regression model, $\hat{Y}_i = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j X_{ij} \quad i = 1, 2, \ldots, n$ (2.36)

***Hypothesis testing in multiple linear regression***

The test for significance of regression is to determine whether there is a linear relationship between the response variable $Y$ and the regression variables $X_1, X_2, \ldots, X_k$.

$H_0$: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$

$H_1$: $\beta_j \neq 0$ for at least one $j$.

Rejection of $H_0$ implies that at least one regression variable contributes to the model. The test procedure involves ANOVA and $F$-test (Table 2.12).

$$F_0 = \frac{\left(\dfrac{SS_R}{k}\right)}{\left(\dfrac{SS_e}{n-k-1}\right)} = \frac{MS_R}{MS_e}$$

Reject $H_0$, if $F_0$ exceeds $F_{\alpha,k,n-k-1}$

**TABLE 2.12** ANOVA for significance of regression

| *Source of variation* | *Sum of square* | *Degree of freedom* | *Mean square* | $F_0$ |
|---|---|---|---|---|
| Regression | $SS_R$ | $k$ | $MS_R$ | $MS_R/MS_e$ |
| Error | $SS_e$ | $n - k - 1$ | $MS_e$ | |
| Total | $SS_{\text{Total}}$ | $n - 1$ | | |

where, $SS_{\text{Total}} = Y^{\text{T}}Y - n\bar{Y}^2$

$SS_R = \beta^{\text{T}}X^{\text{T}}Y - n\bar{Y}^2$

$SS_e = SS_{\text{Total}} - SS_R$

These computations are usually performed with regression software or any other statistical software. The software also reports other useful statistics.

### *Coefficient of multiple determination ($R^2$)*

$R^2$ is a measure of the amount of variation explained by the model.

$$R^2 = \frac{SS_R}{SS_{\text{Total}}} = 1 - \left(\frac{SS_e}{SS_{\text{Total}}}\right) \tag{2.37}$$

A large value of $R^2$ does not necessarily imply that the regression model is a better one. Adding a variable to the model will always increase the value of $R^2$, regardless of whether this additional variable is statistically significant or not. Hence, we prefer to use an adjusted $R^2$.

$$R^2_{\text{adj}} = 1 - \frac{SS_e/n - p}{SS_{\text{Total}}/n - 1} = 1 - \frac{n - 1}{n - p}(1 - R^2) \tag{2.38}$$

where $n - p$ are the error degrees of freedom.

In general $R^2_{\text{adj}}$ value will not increase as variables are added to the model. The addition of unnecessary terms to the model decreases the value of $R^2_{\text{adj}}$.

**ILLUSTRATION 2.2**

**Multiple Linear Regression**

An experiment was conducted to determine the lifting capacity of industrial workers. The worker characteristics such as age, body weight, body height and Job Specific Dynamic Lift Strength (JSDLS) are used as independent variables and Maximum Acceptable Load of Lift (MALL) is the dependant variable. The data collected is given Table 2.13.

**TABLE 2.13** Data for Illustration 2.2

| *MALL* | *Age* | *Height* | *Weight* | *JSDLS* |
|---|---|---|---|---|
| 21.5 | 42 | 174 | 73 | 51 |
| 22.5 | 24 | 174 | 60 | 58 |
| 23.5 | 24 | 166 | 58 | 55 |
| 20.75 | 31 | 159 | 49 | 44 |
| 22.5 | 38 | 160 | 61 | 49 |
| 23.5 | 30 | 162 | 52 | 52 |
| 22.00 | 34 | 165 | 46 | 48 |
| 20.50 | 28 | 167 | 46 | 48 |
| 16.0 | 28 | 164 | 39 | 35 |
| 17.75 | 22 | 161 | 47 | 40 |

When more than two independent variables are involved, we normally use computer software for obtaining solution. The solution obtained from computer software is given as follows:

```
Y = MALL(RESPONSE), X1 = Age, X2 = Height, X3 = Weight, X4 = JSDLS
```

**Statistical Analysis**

The regression equation is

MALL = 28.7 + 0.106 X1 – 0.178 X2 – 0.0365 X3 + 0.427 X4

| *Predictor* | *Coefficient* | *Se of Coef* | *p* |
|---|---|---|---|
| Constant | 28.732 | 4.481 | 0.001 |
| Age(X1) | 0.10567 | 0.02544 | 0.009 |
| Height(X2) | -0.17808 | 0.02959 | 0.002 |
| Weight(X3) | -0.03647 | 0.02324 | 0.177 |
| JSDLS(X4) | 0.42693 | 0.02803 | 0.000 |

R-Sq = 98.7% R-Sq(adj) = 97.6%

**Analysis of Variance**

| *Source* | *SS* | *df* | *MS* | *F* | *P* |
|---|---|---|---|---|---|
| Regression | 53.393 | 4 | 13.348 | 94.67 | 0.000 |
| Error | 0.707 | 5 | 0.141 | | |
| Total | 54.00 | 9 | | | |

**Residuals**

| *Obs* | *X1* | *Y* | $\hat{Y}$ | *Residual* | *Standardized residual* |
|---|---|---|---|---|---|
| 1 | 42.0 | 21.500 | 21.295 | 0.205 | 1.64 |
| 2 | 24.0 | 22.500 | 22.856 | -0.356 | -1.50 |
| 3 | 24.0 | 23.500 | 23.073 | 0.427 | 1.46 |
| 4 | 31.0 | 20.750 | 20.691 | 0.059 | 0.18 |
| 5 | 38.0 | 22.500 | 22.950 | -0.450 | -1.74 |
| 6 | 30.0 | 23.500 | 23.357 | 0.143 | 0.45 |
| 7 | 34.0 | 22.000 | 21.757 | 0.243 | 0.92 |
| 8 | 28.0 | 20.500 | 20.766 | -0.266 | -0.84 |
| 9 | 28.0 | 16.000 | 16.006 | -0.006 | -0.03 |
| 10 | 22.0 | 17.750 | 17.749 | 0.001 | 0.00 |

The *p*-value from ANOVA indicates that the model is significant. However, the regression coefficient for the variable WEIGHT is not significant (*p*-value = 0.177). The R-Sq(adj) value (97.6%) shows that the fitted model is a good fit of the data.

Figures 2.5 and 2.6 show the normal probability plot of standardized residuals and the plot of standardized residuals versus fitted values respectively for Illustration 2.2. From these figures it is inferred that the model is a good fit of the data.

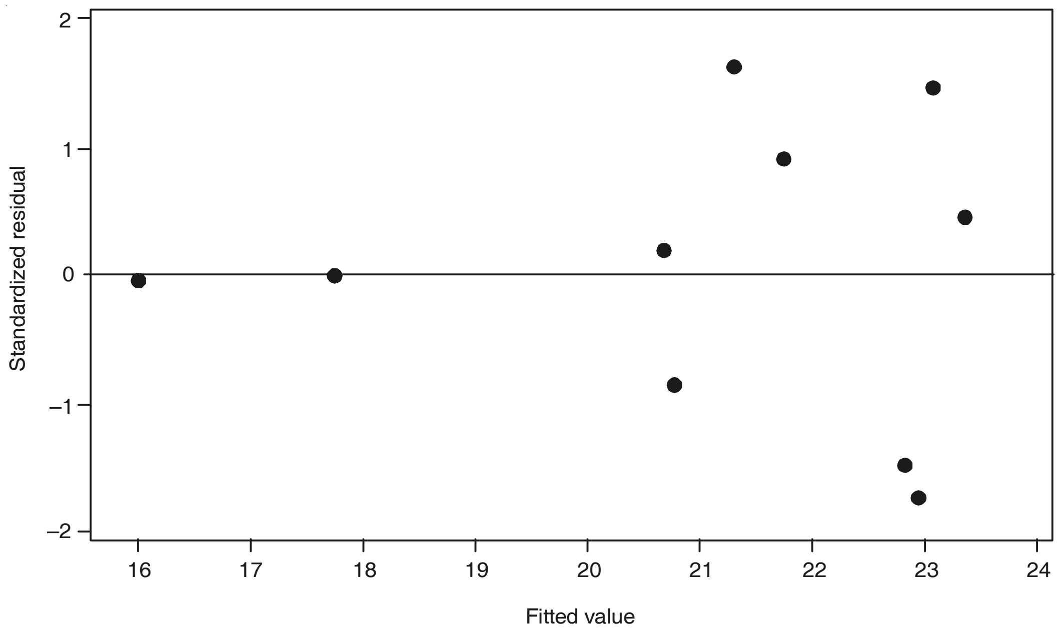



**FIGURE 2.5** Plot of standardized residuals vs. fitted value for Illustration 2.2.

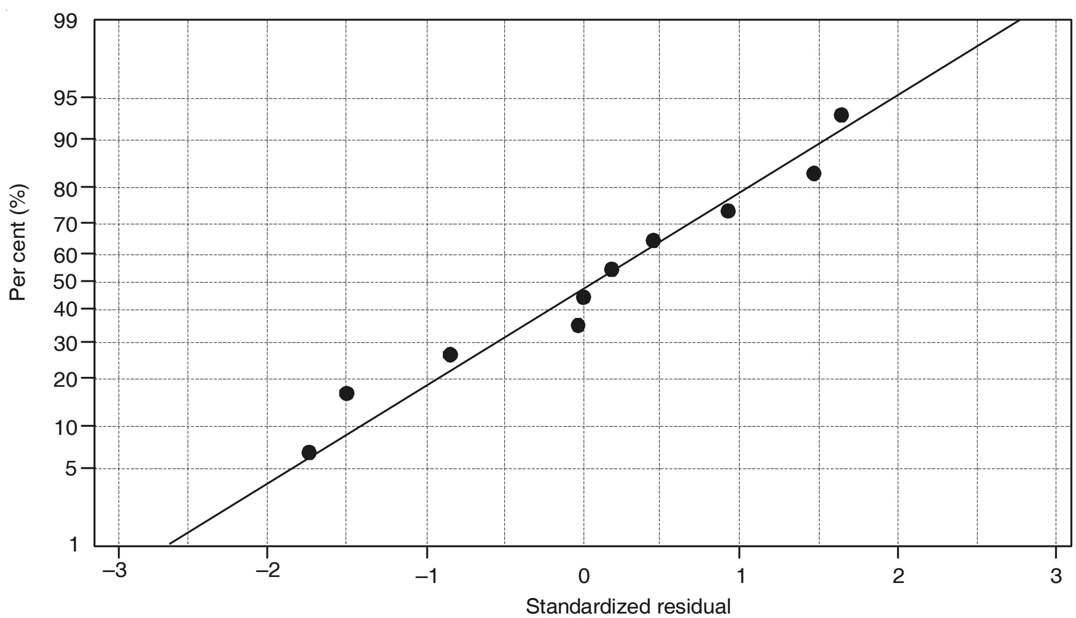



**FIGURE 2.6** Normal probability plot of standardized residuals for Illustration 2.2.

## PROBLEMS

**2.1** In a small town, a hospital is planning for future needs in its maternity ward. The data in Table 2.14 show the number of births in the last eight years.

**TABLE 2.14** Data for Problem 2.1

| Year: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Births: | 565 | 590 | 583 | 597 | 615 | 611 | 610 | 623 |

(i) Develop a simple linear regression model to the data for estimating the number of births.
(ii) Test the significance of regression using $F$-test
(iii) Test significance of $\beta_1$

**2.2** The data in Table 2.15 shows the yield of sugar and the quantity of sugarcane crushed. Develop a simple linear regression equation to estimate the yield.

**TABLE 2.15** Data for Problem 2.2

| *Yield* (tons) | *Quantity crushed* (tons) |
|---|---|
| 100 | 250 |
| 94 | 210 |
| 89 | 195 |
| 95 | 220 |
| 101 | 245 |
| 97 | 196 |
| 88 | 189 |

**2.3** Obtain normal probability plot for the following data. Estimate the mean and standard deviation from the plot. Compare these values with the actual values computed from the data.

Data: 230, 220, 225, 250, 235, 260, 245, 255, 210, 215, 218, 212, 240, 245, 242, 253, 222, 257, 248, 232.

**2.4** An induction hardening process was studied with three parameters namely Power potential ($P$), Scan speed ($S$) and Quenching flow rate ($Q$). The surface hardness in HRA was measured with Rockwell Hardness tester. The data collected is given in Table 2.16.

**TABLE 2.16** Data for Problem 2.4

| *Hardness in HRA* | $P$ | $Q$ | $S$ |
|---|---|---|---|
| 80 | 6.5 | 1.5 | 15 |
| 80 | 6.5 | 1.5 | 17.5 |
| 77 | 6.5 | 1.5 | 20 |
| 77 | 6.5 | 2.0 | 15 |
| 82 | 6.5 | 2.0 | 17.5 |

(*Contd.*)

**TABLE 2.16** Data for Problem 2.4 (*Contd.*)

| *Hardness in HRA* | *P* | *Q* | *S* |
|---|---|---|---|
| 80 | 7.5 | 1.5 | 15 |
| 79 | 7.5 | 1.5 | 2.0 |
| 75 | 7.5 | 1.5 | 20 |
| 66 | 8.5 | 1.5 | 15 |
| 62 | 8.5 | 1.5 | 17.5 |
| 61 | 8.5 | 2.5 | 20 |
| 64 | 8.5 | 2.5 | 17.5 |

(i) Fit a multiple regression model to these data.
(ii) Test for significance of regression. What conclusions will you draw?
(iii) Based on *t*-tests, do we need all the four regressor variables?

**2.5** The average of pulmonary ventilation (PV) obtained from a manual lifting experiment for different work loads is given in Table 2.17.

**TABLE 2.17** Data for Problem 2.5

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Work load: (kg m/min) | 10 | 20 | 30 | 35 | 44 | 54 | 60 | 67 | 75 | 84 | 92 |
| PV: (L/min) | 10.45 | 12.2 | 15.7 | 17.59 | 19.03 | 18.39 | 24.57 | 23.00 | 21.23 | 26.8 | 22.7 |

(i) Develop a simple linear regression model to the data for estimating the pulmonary ventilation (PV).
(ii) Test the significance of regression using *F*-test.
(iii) Test the significance of $\beta_1$.

**2.6** The following factor effects (Table 2.18) have been obtained from a study.
(i) Plot the effects on a normal probability paper and identify significant effects if any.
(ii) Obtain half-normal plot of these effects and identify the significant effects.
(iii) Compare these two plots and give your remarks.

**TABLE 2.18** Factor effects

| | | | |
|---|---|---|---|
| (1) = 7 | ad = 10 | ae = 12 | de = 6 |
| a = 9 | bd = 32 | be = 35 | ade = 10 |
| b = 34 | abd = 50 | abe = 52 | bde = 30 |
| c = 16 | cd = 18 | ce = 15 | abde = 53 |
| d = 8 | acd = 21 | ace = 22 | cde = 15 |
| e = 8 | bcd = 44 | bce = 45 | acde = 20 |
| ab = 55 | abcd = 61 | abce = 65 | bcde = 41 |
| ac = 20 | bc = 40 | abc = 60 | abcde = 63 |

CHAPTER

# 3 Single-factor Experiments

## 3.1 INTRODUCTION

In single-factor experiments only one factor is investigated. The factor may be either qualitative or quantitative. If the levels of a factor are qualitative (type of tool, type of material, etc.), it is called *qualitative factor*. If the levels of a factor are quantitative (temperature, pressure, velocity, etc.), it is called *quantitative factor*. The levels of a factor can be fixed (selecting specific levels) or random (selecting randomly).

Some examples of a single-factor experiment are:

- Studying the effect of type of tool on surface finish of a machined part
- Effect of type of soil on yield
- Effect of type of training program on the performance of participants
- Effect of temperature on the process yield
- Effect of speed on the surface finish of a machined part

If the levels are fixed, the associated statistical model is called *fixed effects model*. Each level of the factor considered to be *treatment*.

## 3.2 COMPLETELY RANDOMIZED DESIGN

In a single-factor experiment if the order of experimentation as well as allocation of experimental units (samples) is completely random, it is called *completely randomized design*.

### 3.2.1 The Statistical Model

$$Y_{ij} = \mu + T_i + e_{ij} \begin{cases} i = 1, 2, ..., a \\ j = 1, 2, ..., n \end{cases} \tag{3.1}$$

where,

$Y_{ij}$ = $j$th observation of the $i$th treatment/level

$\mu$ = overall mean

$T_i$ = $i$th treatment effect
$e_{ij}$ = error

Equation (3.1) is a linear statistical model often called the *effects model*. Also it is referred as one-way or single-factor Analysis of Variance (ANOVA) model. The objective here is to test the appropriate hypotheses about the treatment means and estimate them.

For hypothesis testing, the model errors are assumed to be normally independently distributed random variables with mean zero and variance $\sigma^2$. And $\sigma^2$ is assumed as constant for all levels of the factor.

The appropriate hypotheses are as follows:

$H_0$: $T_1 = T_2 = \ldots = T_a = 0$
$H_1$: $T_i \neq 0$, at least for one $i$

Here we are testing the equality of treatment means or testing that the treatment effects are zero.

The appropriate procedure for testing $a$ treatment means is ANOVA.

### 3.2.2 Typical Data for Single-factor Experiment

The data collected from a single-factor experiment can be shown as in Table 3.1.

**TABLE 3.1** Typical data for single-factor experiment

| *Treatment (level)* | *Observations* | *Total* | *Average* |
|---|---|---|---|
| 1 | $Y_{11}\ Y_{12}\ \ldots\ Y_{1n}$ | $T_{1.}$ | $\bar{Y}_{1.}$ |
| 2 | $Y_{21}\ Y_{22}\ \ldots\ Y_{2n}$ | $T_{2.}$ | $\bar{Y}_{2.}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $a$ | $Y_{a1}\ Y_{a2}\ \ldots\ Y_{an}$ | $T_{a.}$ | $\bar{Y}_{a.}$ |
| | Total | $T_{..}$ | $\bar{Y}_{..}$ |

The dot (.) indicates the summation over that subscript.
$Y_{ij}$ represents the $j$th observation under the factor level or treatment $i$.

Generally, there will be $n$ observations under $i$th treatment.

$T_{i.}$ represents the total of the observations under the $i$th treatment.

$\bar{Y}_{i.}$ is the average of the $i$th treatment.

$T_{..}$ is the grand total.

$\bar{Y}_{..}$ is the grand average.

$$T_{i.} = \sum_{j=1}^{n} Y_{ij} \tag{3.2}$$

$$\bar{Y}_{i.} = \sum_{i=1}^{a} \frac{Y_{i.}}{n} \qquad i = 1, 2, \ldots, a \tag{3.3}$$

$$Y_{..} = \sum_{i=1}^{a} \sum_{j=1}^{n} Y_{ij} \tag{3.4}$$

$$\bar{Y}_{..} = \frac{Y_{..}}{N} \tag{3.5}$$

where $N = an$, the total number of observations.

### 3.2.3 Analysis of Variance

As already mentioned, the name ANOVA is derived from partitioning of total variability into its component parts.

The total corrected Sum of Squares ($SS$)

$$SS_{\text{Total}} = \sum_{i=1}^{a} \sum_{j=1}^{n} (Y_{ij} - \bar{Y}_{..})^2 \tag{3.6}$$

is used as a measure of overall variability in the data. It is partitioned into two components.

Total variation = variation between treatments + variation within the treatment or error

$SS_{\text{Total}} = SS$ due to treatments + $SS$ due to error

i.e., $SS_{\text{Total}} = SS_T + SS_e$

A typical format used for ANOVA computations is shown in Table 3.2

**TABLE 3.2** ANOVA: Single-factor experiment

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ |
|---|---|---|---|---|
| Between treatments | $SS_T$ | $a - 1$ | $MS_T = \dfrac{SS_T}{a-1}$ | $\dfrac{MS_T}{MS_e}$ |
| Within treatments (error) | $SS_e$ | $N - a$ | $MS_e = \dfrac{SS_e}{N-a}$ | |
| Total | $SS_{\text{Total}}$ | $N - 1$ | | |

Mean square is also called *variance*.

If $F_0 > F_{\alpha,a-1,N-a}$, Reject $H_0$.

Generally, we use 5% level of significance ($\alpha = 5\%$) for testing the hypothesis in ANOVA. $P$-value can also be used.

### 3.2.4 Computation of Sum of Squares

Let $T_{..}$ = grand total of all observations/response ($Y_{..}$)
$N$ = total number of observations
$n$ = number of replications/number of observations under the $i$th treatment
$SS$ = sum of squares
CF = correction factor
$T_{i.}$ = $i$th treatment total

$$\text{CF} = \frac{T_{..}^2}{N} \tag{3.7}$$

$$SS_{\text{Total}} = \sum_{i=1}^{a} \sum_{j=1}^{n} Y_{ij}^2 - \text{CF} \tag{3.8}$$

$$SS_T = \sum_{i=1}^{a} \frac{T_{i.}^2}{n} - \text{CF} \tag{3.9}$$

$$SS_e = SS_{\text{Total}} - SS_T \tag{3.10}$$

Note that the factor level (treatment) totals are used to compute the treatment (factor) sum of squares.

**ILLUSTRATION 3.1**

**Completely Randomized Design**

A manufacturing engineer wants to investigate the effect of feed rate (mm/min) on the surface finish of a milling operation. He has selected three different feed rates, i.e., 2, 4 and 6 mm/min for study and decided to obtain four observations at each feed rate. Thus, this study consists of 12 experiments (3 levels × 4 observations). Since the order of experimentation should be random, a test sheet has to be prepared as explained below. The 12 experiments are serially listed in Table 3.3.

**TABLE 3.3** Experiment run number

| *Feed rate* (mm/min) | *Observations* (*Surface roughness*) | | | |
|---|---|---|---|---|
| 2 | 1 | 2 | 3 | 4 |
| 4 | 5 | 6 | 7 | 8 |
| 6 | 9 | 10 | 11 | 12 |

Now, by selecting a random number between 1 and 12, we can determine which experiment should be run first. Suppose the first random number is 7. Then the number 7th experiment (observation) with feed rate 4 mm/min is run. This procedure is repeated till all the experiments are scheduled randomly. Table 3.4 gives the test sheet with the order of experimentation randomized. The randomization of test order is required to even out the effect of extraneous variables (nuisance variables).

**TABLE 3.4** Test sheet for Illustration 3.1

| *Order of experiment (time sequence)* | *Run number* | *Feed rate* |
|---|---|---|
| 1 | 7 | 4 |
| 2 | 3 | 2 |
| 3 | 11 | 6 |
| 4 | 2 | 2 |
| 5 | 6 | 4 |
| 6 | 1 | 2 |
| 7 | 5 | 4 |
| 8 | 9 | 6 |
| 9 | 4 | 2 |
| 10 | 10 | 6 |
| 11 | 8 | 4 |
| 12 | 12 | 6 |

Suppose that the engineer has conducted the experiments in the random order given in Table 3.4. The data collected on the surface roughness are given in Table 3.5.

**TABLE 3.5** Data from surface finish experiment of Illustration 3.1

| *Feed rate* (mm/min) | *Observations* (*Surface roughness*) | | | | $T_{i.}$ | $\bar{Y}_i$ |
|---|---|---|---|---|---|---|
| 2 | 7.0 | 7.5 | 7.8 | 8.3 | 30.6 | 7.65 |
| 4 | 5.8 | 4.6 | 4.8 | 6.2 | 21.4 | 5.35 |
| 6 | 9.2 | 9.6 | 8.2 | 8.5 | 35.5 | 8.88 |

$T_{..}$ = 87.5
$\bar{Y}_{..}$ = 7.29

The data can be analysed through analysis of variance. The hypothesis to be tested is $H_0$: $\mu_1 = \mu_2 = \mu_3$ against $H_1$: at least one mean is different.

Let $T_{i.}$ = treatment totals
$T_{..}$ = grand total

These totals are calculated in Table 3.5.

**Computation of sum of squares**

$$\text{Correction factor (CF)} = \frac{T_{..}^2}{N} \quad \text{(from Eq. 3.7)}$$

$$= \frac{(87.5)^2}{12} = 638.021$$

$$SS_{\text{Total}} = \sum_{i=1}^{a}\sum_{j=1}^{n} Y_{ij}^2 - \text{CF} \quad \text{(from Eq. 3.8)}$$

$$= \{(7)^2 + (7.5)^2 + \cdots + (8.5)^2\} - \text{CF}$$

$$= 667.550 - 638.021$$

$$= 29.529$$

$$SS_T = \sum_{i=1}^{a} \frac{T_{i.}^2}{n_i} - \text{CF} \quad \text{(from Eq. 3.9)}$$

$$= \frac{(30.6)^2}{4} + \frac{(21.4)^2}{4} + \frac{(37.5)^2}{4} - \text{CF}$$

$$= \frac{(30.6)^2 + (21.4)^2 + (37.5)^2}{12} - 638.021$$

$$= 663.643 - 638.021 = 25.622$$

The analysis of variance is given in Table 3.6. The error sum of squares is obtained by substraction.

**TABLE 3.6** ANOVA for the surface finish experiment (Illustration 3.1)

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ |
|---|---|---|---|---|
| Feed rate | 25.622 | 2 | 12.811 | 29.52 |
| Error | 3.907 | 9 | 0.434 | |
| Total | 29.529 | 11 | | |

$F_{5\%,2,9} = 4.26$

Since $F_0 = 29.52 > 4.26$, $H_0$ is rejected.

The inference is that the treatment means are significantly different at 5% level of significance. That is, the treatment (feed rate) has significant effect on the surface finish.

### 3.2.5 Effect of Coding the Observations

Sometimes, the magnitude of the response may be a larger value involving three or four digits. In such cases the resultant sum of squares would be a very large value. To simplify the computations, we adopt the coding of the observations (Response). For this purpose, we select an appropriate value and subtract from each individual observation and obtain coded observations. Using these data, we compute the sum of squares and carryout ANOVA. These results will be same as those obtained on the original data. However, for determining the optimal level, it is preferable to use the mean values on original data.

For illustration purpose, let us consider the experiment on surface finish (Table 3.5). Subtracting 7.0 from each observation, we obtain the coded data as given in Table 3.7.

**TABLE 3.7** Coded data for the surface finish experiment (Illustration 3.1)

| *Feed rate* (mm/min) | *Observations* (*Surface roughness*) | | | | $T_{i.}$ |
|---|---|---|---|---|---|
| 2 | 0.0 | 0.5 | 0.8 | 1.3 | 2.6 |
| 4 | –1.2 | –2.4 | –2.2 | – 0.8 | – 6.6 |
| 6 | 2.2 | 2.6 | 1.2 | 1.5 | 7.5 |

$T_{..} = 3.5$

**Computation of sum of squares**

$$\text{CF} = \frac{T_{..}^2}{N} = \frac{(3.5)^2}{12} = 1.021$$

$$SS_{\text{Total}} = \sum_{i=1}^{a}\sum_{j=1}^{n} Y_{ij}^2 - \text{CF}$$

$$= \{(0.0)^2 + (0.5)^2 + \cdots + (1.5)^2\} - \text{CF}$$

$$= 30.550 - 1.021$$

$$= 29.529$$

$$SS_T = \sum_{i=1}^{a} \frac{T_{i.}^2}{n_i} - \text{CF}$$

$$= \frac{(2.6)^2}{4} + \frac{-(6.6)^2}{4} + \frac{(7.5)^2}{4} - \text{CF}$$

$$= 26.643 - 1.021 = 25.622$$

These sum of square computations are same as the one obtained with the original data. From this, it is evident that coding the observations by subtracting a constant value from all the observations will produce the same results.

### 3.2.6 Estimation of Model Parameters

The single-factor model is

$$Y_{ij} = \mu + T_i + e_{ij} \tag{3.11}$$

It is a linear statistical model with parameters $\mu$ and $T_i$.

$$\hat{\mu} = \overline{Y}_{..} \tag{3.12}$$

$$\hat{T}_i = \overline{Y}_{i.} - \overline{Y}_{..}, \quad i = 1, 2, \ldots, a \tag{3.13}$$

That is, the overall mean $\mu$ is estimated by the grand average and the treatment effect is equal to the difference between the treatment mean and the grand mean.

The error $\hat{e}_{ij} = Y_{ij} - \mu - T_i$

$$= Y_{ij} - \overline{Y}_{..} - (\overline{Y}_{i.} - \overline{Y}_{..})$$

$$= Y_{ij} - \overline{Y}_{i.} \tag{3.14}$$

The confidence interval for the $i$th treatment mean is given by

$$\overline{Y}_{i.} - t_{\alpha/2, N-a}\sqrt{\frac{MS_e}{n}} \le \mu_i \le \overline{Y}_{i.} + t_{\alpha/2, N-a}\sqrt{\frac{MS_e}{n}} \tag{3.15}$$

And the confidence interval on the difference in any two treatment means, say $(\mu_i - \mu_j)$ is given by

$$(\overline{Y}_{i.} - \overline{Y}_{.j}) - t_{\alpha/2, N-a}\sqrt{\frac{MS_e}{n}} \le (\mu_i - \mu_j) \le (\overline{Y}_{i.} - \overline{Y}_{.j}) + t_{\alpha/2, N-a}\sqrt{\frac{MS_e}{n}} \tag{3.16}$$

For Illustration 3.1, the 95% confidence interval for the treatment 2 can be computed using Eq. (3.15) as follows:

$$\hat{\mu} = \frac{87.5}{12} = 7.29$$

$$\hat{T}_2 = \overline{Y}_{2.} - \overline{Y}_{..}$$

$$= 5.35 - 7.29$$

$$= -1.94$$

At 5% significance level $t_{0.025,9} = 2.262$

The 95% confidence interval for the treatment 2 is

$$5.35 - 2.262\sqrt{\frac{0.434}{4}} \le \mu_2 \le 5.35 - 2.262\sqrt{\frac{0.434}{4}}$$

$$5.35 - 0.745 \le \mu_2 \le 5.35 - 0.745$$

$$4.605 \le \mu_2 \le 6.095$$

### 3.2.7 Model Validation

The ANOVA equation is an algebraic relationship. The total variance is partitioned to test that there is no difference in treatment means. This requires the following assumptions to be satisfied:

1. The observations are adequately described by the model

$$Y_{ij} = \mu + T_i + e_{ij}$$

2. The errors are independent and normally distributed.
3. The errors have constant variance $\sigma^2$.

If these assumptions are valid, the ANOVA test is valid. This is carried out by analysing the residuals ($R_{ij}$).

The residual for the $j$th observation of the $i$th treatment is given by

$$R_{ij} = Y_{ij} - \hat{Y}_{ij} \text{ (prediction error)} \tag{3.17}$$

where, $\hat{Y}_{ij} = \hat{\mu} + \hat{T}_i$

$$= \bar{Y}_{..} + (\bar{Y}_{i.} - \bar{Y}_{..})$$

$$= \bar{Y}_{i.} \tag{3.18}$$

Therefore,
$$R_{ij} = Y_{ij} - \bar{Y}_{i.} \tag{3.19}$$

That is, the estimate of any observation in the $i$th treatment is simply the corresponding treatment average. That is, $R_{ij} = Y_{ij} - \bar{Y}_{i.}$. If the model is adequate, the residuals must be structure less. The following graphical analysis is done to check the model.

### *Estimation of residuals*

Data and residuals of Illustration 3.1 are given in Table 3.8.

$$\text{Residual } (R_{ij}) = Y_{ij} - \bar{Y}_{i.} \text{ (Eq. 3.19)}$$

**TABLE 3.8** Residuals for the surface finish experiment (Illustration 3.1)

| *Feed rate* | *Observations (Surface roughness)* | | | | $\hat{Y}_{ij} = \bar{Y}_{i.}$ |
|---|---|---|---|---|---|
| 2 | 7.0<br>– 0.65 | 7.5<br>– 0.15 | 7.8<br>0.15 | 8.3<br>0.65 | 7.65 |
| 4 | 5.8<br>0.45 | 4.6<br>– 0.75 | 4.8<br>– 0.55 | 6.2<br>0.85 | 5.35 |
| 6 | 9.2<br>0.32 | 9.6<br>0.72 | 8.2<br>– 0.68 | 8.5<br>– 0.38 | 8.88 |

The residuals are given in the left corner of each cell in Table 3.8.

### *Check to determine the outliers*

An outlier is an observation obtained from the experiment which is considered as an *abnormal response*, which should be removed from the data before the data are analysed. Any residual which is away from the straight line passing through the residual plot on a normal probability paper is an outlier. These can also be identified by examining the standardized residuals ($z_{ij}$)

$$z_{ij} = \frac{R_{ij}}{\sqrt{MS_e}} \tag{3.20}$$

If the errors are NID (0, $\sigma^2$), the standardized residuals follow normal distribution with mean zero and variance one. Thus, about 95% of residuals should fall within ± 2 and all of them (99.73%) within ± 3. A residual bigger than 3 or 4 standard deviations from zero, can be considered as an *outlier*. From Illustration 3.1, the largest standard residual is

$$\frac{R_{ij}}{\sqrt{MS_e}} = \frac{0.85}{\sqrt{0.434}} = 1.29$$

This indicates that all residuals are less than ±2 standard deviations. Hence, there are no outliers. This can be verified in the normal plot also.

### Model adequacy checking

**1.** ***Check for normality:*** In this assumption errors are normally distributed. The plot of residuals on a normal probability paper shall resemble a straight line for the assumption to be valid. Figure 3.1 shows the plot of residuals on the normal probability paper for Illustration 3.1. The plot has been obtained as explained in Chapter 2. From the figure it can be inferred that the errors are normally distributed. Moderate departures from the straight line may not be serious, since the sample size is small. Also observe that there are no outliers.

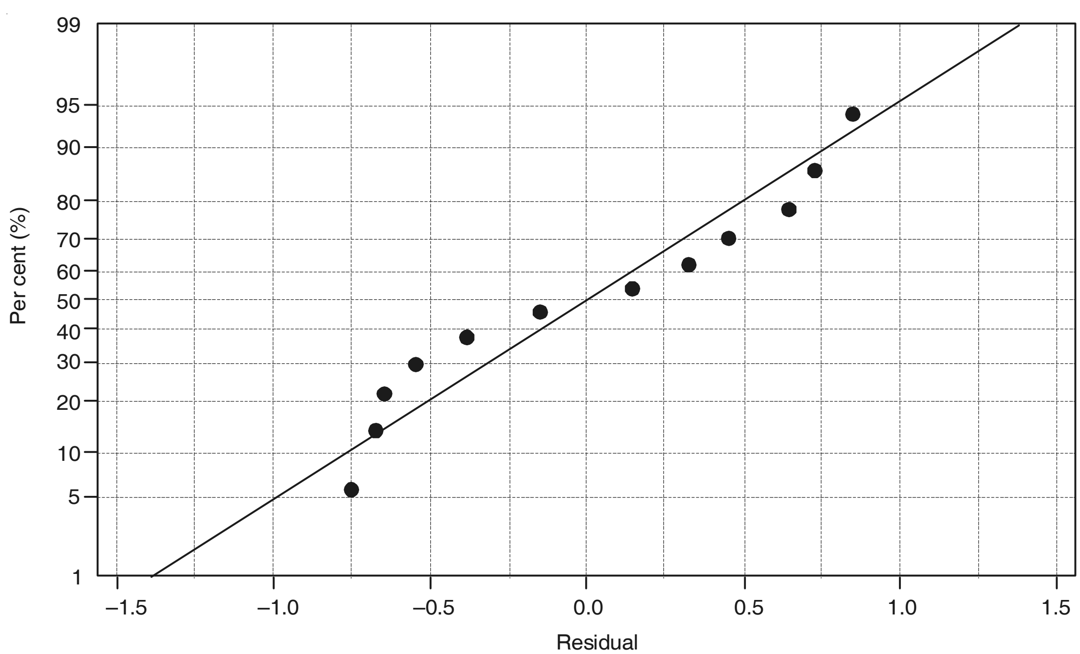



**FIGURE 3.1** Normal probability plot of residuals for Illustration 3.1.

**2.** ***Check for the independence:*** If the plot of errors (residuals) in the order of data collection shows a random scatter, we infer that the errors are independent. If any correlation exists between the residuals, it indicates the violation of the independent assumption. Table 3.9 gives the residuals as per the order of experimentation (Random order). Figure 3.2 shows the plot of residuals versus order of experimentation. From Figure 3.2, it can be seen that the residuals are randomly scattered indicating that they are independent.

**TABLE 3.9** Residuals as per the order of experimentation for Illustration 3.1

| *Order of experiment (time sequence)* | *Run number* | *Feed rate* | *Residual* |
|---|---|---|---|
| 1 | 7 | 4 | – 0.55 |
| 2 | 3 | 2 | 0.15 |
| 3 | 11 | 6 | – 0.68 |
| 4 | 2 | 2 | – 0.15 |
| 5 | 6 | 4 | – 0.75 |
| 6 | 1 | 2 | – 0.65 |
| 7 | 5 | 4 | 0.45 |
| 8 | 9 | 6 | 0.32 |
| 9 | 4 | 2 | 0.65 |
| 10 | 10 | 6 | 0.72 |
| 11 | 8 | 4 | 0.85 |
| 12 | 12 | 6 | – 0.38 |

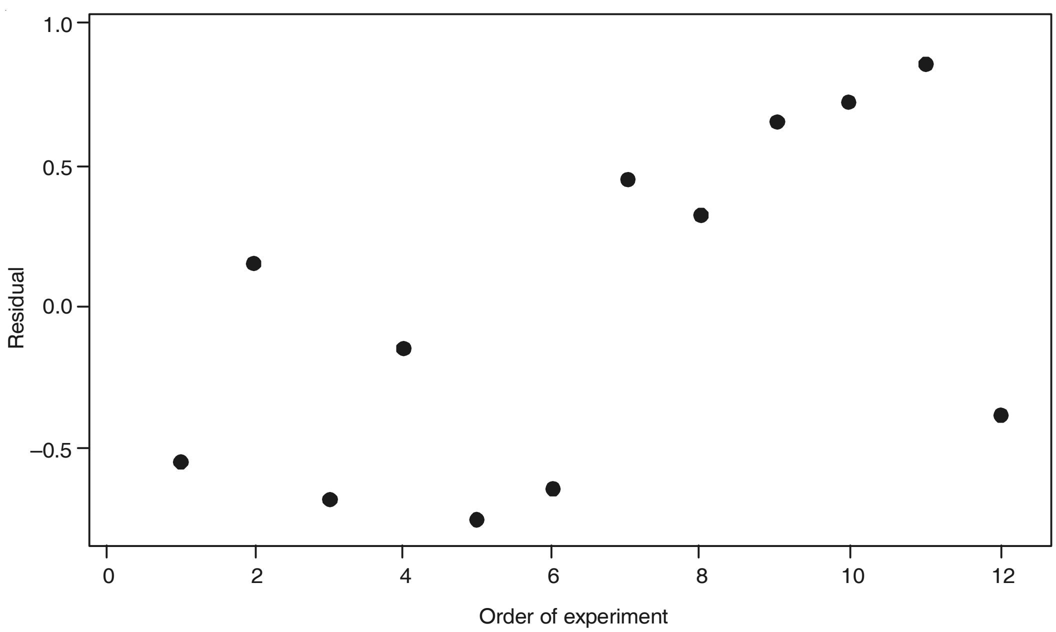



**FIGURE 3.2** Plot of residuals vs. order of experimentation for Illustration 3.1.

**3. *Check for constant variance:*** If the model is adequate, the plot of residuals should be structure less and should not be related to any variable including the predicted response ($\hat{Y}_{ij}$). The plots of residuals versus predicted response should appear as a parallel band centered about zero. If the spread of residuals increases as $\hat{Y}_{ij}$ increases, we can infer that the error variance increases with the mean. Figure 3.3 shows the plot of residuals versus $\hat{Y}_{ij}$ for Illustration 3.1.

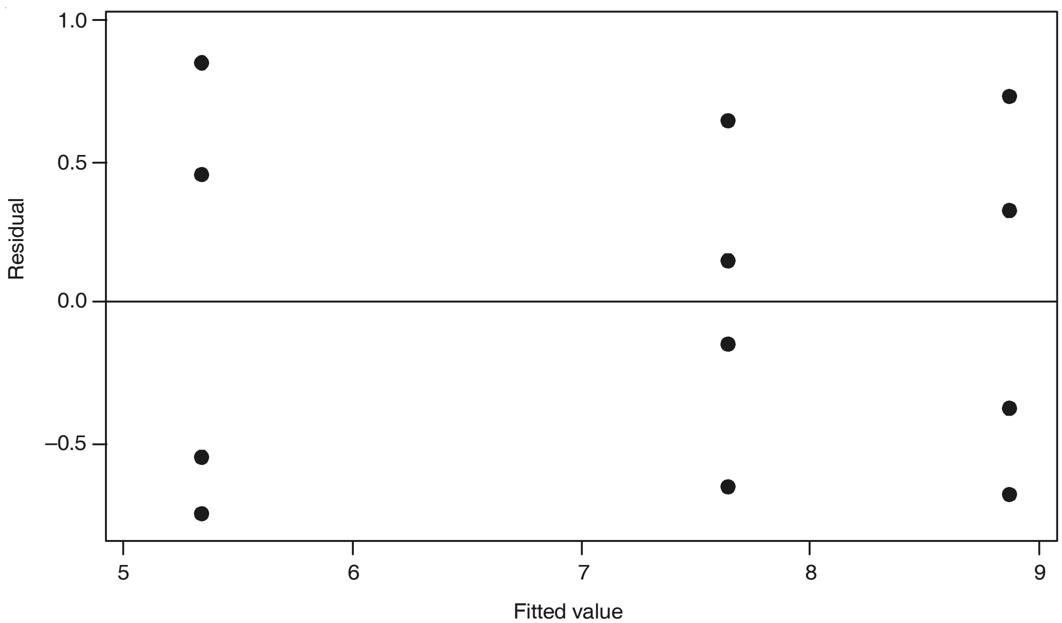



**FIGURE 3.3** Plot of residuals vs. fitted values for Illustration 3.1.

From Figure 3.3, it is observed that no unusual structure is present indicating that the model is correct.

### 3.2.8 Analysis of Treatment Means

Suppose in a single-factor experiment, the ANOVA reveals the rejection of null hypothesis. That is, there are differences between the treatment means, but which means differ is not known. Hence, the comparison among these treatment means may be useful. This is also useful to identify the best and preferred treatments for possible use in practice. The best treatment is the one that corresponds to the treatment mean which optimizes the response. Other preferred treatments can be identified through multiple comparison methods. These tests are employed after analysis of variance is conducted. The following are some of the important pair wise comparison methods:

- Duncan’s multiple range test
- Newman–Keuls test
- Fisher’s Least Significant Difference (LSD) test and
- Turkey’s test

### *Duncan's multiple range test*

This test is widely used for comparing all pairs of means. The comparison is done in a specific manner. The following are the steps:

*Step* 1: Arrange the $a$ treatment means in ascending order.

*Step* 2: Compute the standard error ($S_e$) of mean $(S_{\bar{Y}_i}) = \sqrt{\dfrac{MS_e}{n}}$ (3.21)

If $n$ is different for the treatments, replace $n$ by $n_h$, where $n_h = \dfrac{a}{\sum_{i=1}^{a} 1/n_i}$

*Step* 3: Obtain values of $r_\alpha(p, f)$ for $p = 2, 3, \ldots, a$, from Duncan's multiple ranges (Appendix A.5).

$\alpha$ = significance level

$f$ = degrees of freedom of error

*Step* 4: Compute least significant ranges ($R_p$) for $p = 2, 3, \ldots, a$.

$$R_p = (S_{\bar{Y}})\ r_\alpha(p, f) \text{ for } p = 2, 3, \ldots, a \tag{3.22}$$

*Step* 5: Test the observed differences between the means against the least significant ranges ($R_p$) as follows:

**Cycle 1:**

- Compare the difference between largest mean and smallest mean with $R_a$
- Compare the difference between largest mean and next smallest mean with $R_{a-1}$ and so on until all comparisons with largest mean are over.

**Cycle 2:**

- Test the difference between the second largest mean and the smallest mean with $R_{a-1}$
- Continue until all possible pairs of means $a\ (a-1)/2$ are tested.

**Inference:** If the observed difference between any two means exceed the least significant range, the difference is considered as significant.

### *Newman–Keuls test*

Operationally, the procedure of this test is similar to that of Duncan's multiple range test except that we use studentized ranges. That is, in Step 4 of Duncan's procedure $R_p$ is replaced by $K_p$, whence

$$K_p \text{ is } q_\alpha(p, f)\, S_{\bar{Y}};\ p = 2, 3, \ldots, a \tag{3.23}$$

where $q_\alpha(p, f)$ values are obtained from studentized range table (Appendix A.4).

### *Fisher's Least Significant Difference (LSD) test*

In this test we compare all possible pairs of means with LSD, where

$$\text{LSD} = t_{\alpha/2,\, N-a} \sqrt{MS_e\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \tag{3.24}$$

If the absolute difference between any two means exceeds LSD, these two means are considered as significantly different. That is, if $|\bar{Y}_{i.} - \bar{Y}_{j.}| > \text{LSD}$, we conclude that the two population means $\mu_i$ and $\mu_j$ differ significantly.

In a balanced design, $n_1 = n_2 = \cdots = n$. Therefore,

$$\text{LSD} = t_{\alpha/2,\, N-a} \sqrt{\frac{2\, MS_e}{n}} \tag{3.25}$$

### *Turkey's test*

In this test we use studentized range statistic. All possible pairs of means are compared with $T_\alpha$,

$$T_\alpha = q_\alpha(a, f) \sqrt{\frac{2\, MS_e}{n}} \tag{3.26}$$

where,

$a$ = number of treatments/levels
$f$ = degrees of freedom associated with $MS_e$

If the absolute difference between any two means exceed $T_\alpha$, we conclude that the two means differ significantly. That is, if $|\bar{Y}_{i.} - \bar{Y}_{j.}| > T_\alpha$, we say that the population means $\mu_i$ and $\mu_j$ differ significantly.

When sample sizes are unequal,

$$T_\alpha = \frac{q_\alpha(a,f)}{\sqrt{2}} \sqrt{MS_e\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \tag{3.27}$$

We can also construct a set of 100(1 – $\alpha$)% confidence intervals for all pairs of means in this procedure as follows:

$$(\bar{Y}_{i.} - \bar{Y}_{.j}) - q_\alpha(a,f)\sqrt{\frac{MS_e}{n}} \le \mu_i - \mu_j \le (\bar{Y}_{i.} - \bar{Y}_{.j}) + q_\alpha(a,f)\sqrt{\frac{MS_e}{n}},\ i \ne j \tag{3.28}$$

**Which pair-wise comparison method to use?**

There are no clear cut rules available to answer this question. Based on studies made by some people, the following guidelines are suggested:

1. Fisher's LSD test is very effective if the $F$-test in ANOVA shows significance at $a = 5\%$.
2. Duncan's multiple range test also give good results.
3. Since Turkey's method controls the overall error rate, many prefer to use it.
4. The power of Newman–Keul's test is lower than that of Duncan's multiple range test and thus it is more conservative.

Illustration 3.2 discusses the above pair wise comparison tests.

## ILLUSTRATION 3.2

**Brick Strength Experiment**

A brick manufacturer had several complaints from his customers about the breakage of bricks. He suspects that it may be due to the clay used. He procures the clay from four different sources ($A$, $B$, $C$ and $D$). He wants to determine the best source that results in better breaking strength of bricks. To study this problem a single-factor experiment was designed with the four sources as the four levels of the factor source. Five samples of bricks have been made with the clay from each source. The breaking strength in kg/cm$^2$ measured is given in Table 3.10.

**TABLE 3.10** Data on brick strength for Illustration 3.2

| *Source* | *Observations* | | | | | $T_{i.}$ | $\bar{Y}_{i.}$ |
|---|---|---|---|---|---|---|---|
| $A$ | 96 | 89 | 94 | 94 | 98 | 471 | 94.2 |
| $B$ | 82 | 84 | 87 | 88 | 84 | 425 | 85 |
| $C$ | 88 | 89 | 92 | 84 | 86 | 439 | 87.8 |
| $D$ | 75 | 79 | 82 | 78 | 76 | 390 | 78 |
| | | | | | | $T_{..} = 1725$ | |

***Data analysis:*** The computations required for ANOVA are performed as follows:

$$\text{CF} = \frac{(1725)^2}{20} = 148781.25$$

$$SS_{\text{Total}} = \sum_{i=1}^{a}\sum_{j=1}^{n} Y_{ij}^2 - \text{CF}$$

$$= 149593 - 148781.25 = 811.75$$

$$SS_T = \frac{\sum_{i=1}^{a} T_{i.}^2}{n} - \text{CF}$$

$$= \frac{(471)^2 + (425)^2 + (439)^2 + (390)^2}{5} - \text{CF}$$

$$= 676.15$$

$$SS_e = SS_{\text{Total}} - SS_T$$

$$= 811.75 - 676.15 = 135.6$$

The computations are summarized in Table 3.11.

The hypotheses to be tested are

$H_0$: $T_1 = T_2 = T_3 = T_4$

$H_1$: $T_i \neq 0$, at least for one $i$

**TABLE 3.11** ANOVA for the brick strength experiment (Illustration 3.2)

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ |
|---|---|---|---|---|
| Between treatments | 676.15 | 3 | 225.38 | 26.59 |
| Error | 135.60 | 16 | 8.475 | |
| Total | 811.75 | 19 | | |

$F_{0.05,3,16} = 3.24$

Since $F_0 > F_\alpha$, we reject $H_0$. That is, the four treatments differ significantly. This indicates that the clay has significant effect on the breaking strength of bricks.

Figures 3.4 and 3.5 show the normal plot of residuals and residuals versus fitted values respectively.

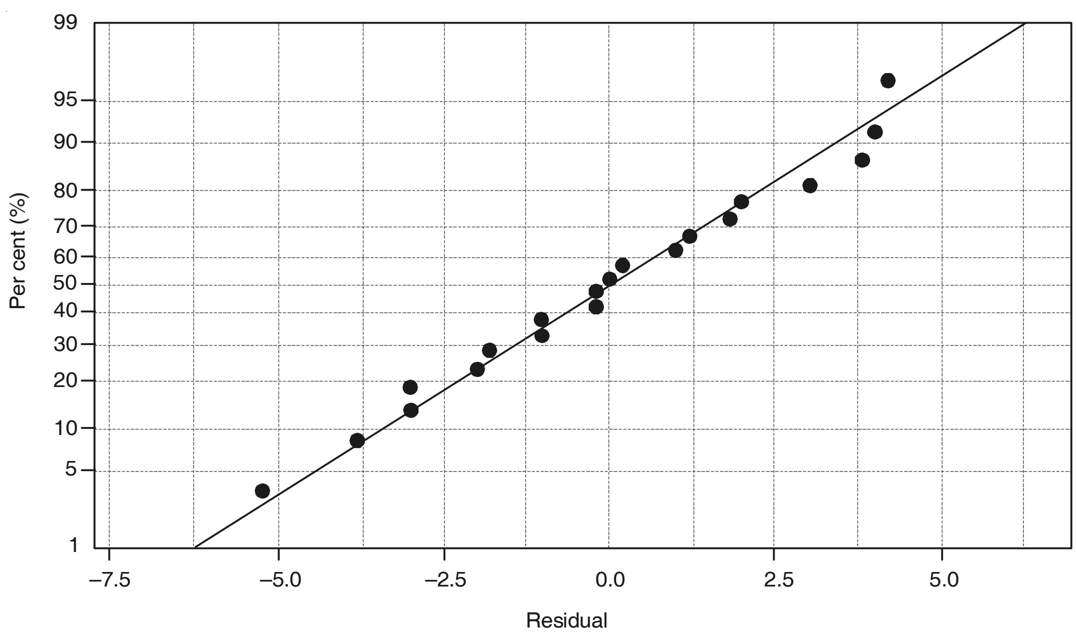



**FIGURE 3.4** Normal probability of residuals for Illustration 3.2.

***Duncan's multiple range test***

*Step* 1: Arrange the treatment means in the ascending order.

| Source | *D* | *B* | *C* | *A* |
|---|---|---|---|---|
| Mean | 78 | 85 | 87.8 | 94.2 |

*Step* 2: Compute the standard error of mean.

$$S_e = \sqrt{\frac{MS_e}{n}} = \sqrt{\frac{8.475}{5}} = 1.30$$

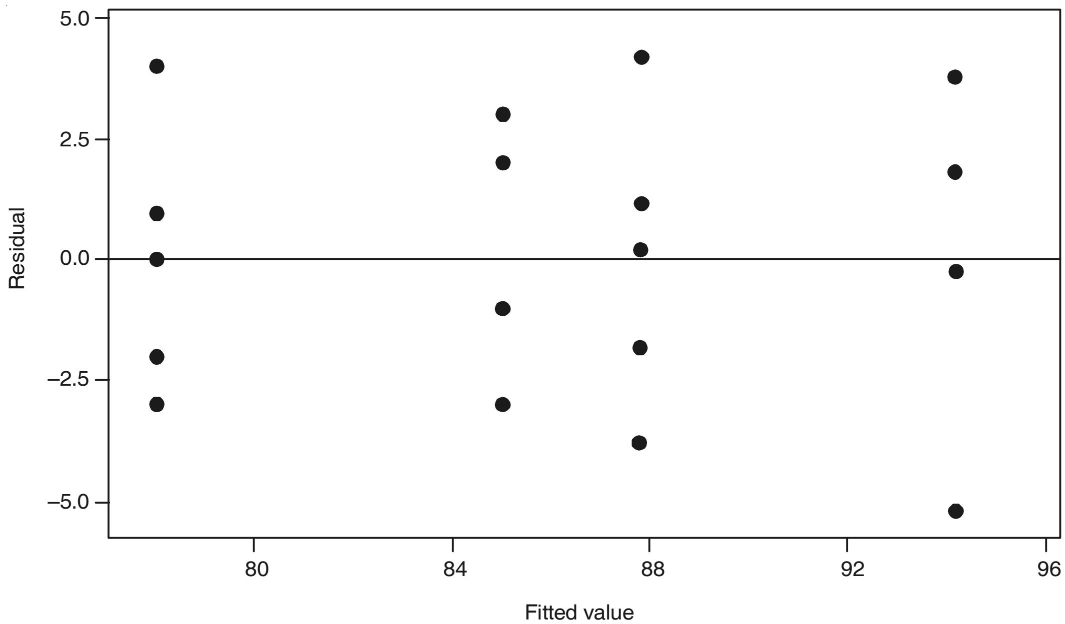



**FIGURE 3.5** Plot of residuals vs. fitted values for Illustration 3.2.

*Step* 3: Select Duncan's multiple ranges from Appendix A.5
For $r_\alpha (p, f)$, $p = 2, 3, 4$, the ranges are:

$$r_{0.05}(p, 16): \quad 3.00 \qquad 3.15 \qquad 3.23$$

*Step* 4: Compute LSR.
The Least Significant Range (LSR) = $r_\alpha(p, 16) \times S_e$
The three Least Significant Ranges are given as follows:

| $R_2$ | $R_3$ | $R_4$ |
|---|---|---|
| 3.9 | 4.095 | 4.199 |

*Step* 5: Compare the pairs of means.
The pair wise comparisons are as follows. The absolute difference is compared with LSR.

**Cycle 1:**

$A$ vs $D$: 94.2 – 78.0 = 16.2 > 4.199 Significant
$A$ vs $B$: 94.2 – 85.0 = 9.2 > 4.095 Significant
$A$ vs $C$: 94.2 – 87.8 = 6.4 > 3.90 Significant

**Cycle 2:**

$C$ vs $D$: 87.8 – 78.0 = 9.8 > 4.095 Significant
$C$ vs $B$: 87.8 – 85.0 = 2.8 < 3.90 Not Significant

**Cycle 3:**

$B$ vs $D$: 85.0 – 78.0 = 7 > 3.90 Significant

We see that there is no significant difference between $B$ and $C$. And $A$ and $D$ differ significantly from $B$ and $C$. These are grouped as follows:

$$\underline{A} \qquad \underline{BC} \qquad \underline{D}$$

***Newman–Keul's test***

*Step* 1:

| Source | $D$ | $B$ | $C$ | $A$ |
|---|---|---|---|---|
| Mean | 78 | 85 | 87.8 | 94.2 |

*Step* 2:

$$S_e = \sqrt{\frac{MS_e}{n}} = \sqrt{\frac{8.475}{5}} = 1.30$$

*Step* 3:

From studentized ranges (Appendix A.4), for $q_\alpha(p, f)$, $p = 2, 3, 4$, the ranges are:

$$q_{0.05}(p,\ 16)\text{: } 3.00 \ 3.65 \ 4.05$$

*Step* 4:

The Least Significant Range (LSR) = $q_\alpha(p,\ 16) \times S_e$

The three Least Significant Ranges are as follows:

| $R_2$ | $R_3$ | $R_4$ |
|---|---|---|
| 3.9 | 4.745 | 5.625 |

*Step* 5:

The pair-wise comparisons are as follows. The absolute difference is compared with LSR.

**Cycle 1:**

$A$ vs $D$: $94.2 - 78.0 = 16.2 > 5.625$ Significant
$A$ vs $B$: $94.2 - 85.0 = 9.2 > 4.745$ Significant
$A$ vs $C$: $94.2 - 87.8 = 6.4 > 3.90$ Significant

**Cycle 2:**

$C$ vs $D$: $87.8 - 78.0 = 9.8 > 4.745$ Significant
$C$ vs $B$: $87.8 - 85.0 = 2.8 < 3.90$ Not significant

**Cycle 3:**

$B$ vs $D$: $85.0 - 78.0 = 7 > 3.90$ Significant

We see that the results of this test are similar to Duncan's multiple range test.

$$\underline{A} \qquad \underline{BC} \qquad \underline{D}$$

**Fisher's Least Significant Difference (LSD) test**

LSD: $t_{\alpha/2,\ N-a}\sqrt{\frac{2MS_e}{n}}$ where $t_{0.025,16} = 2.12$ (Appendix A.2)

$$= 2.12 \times \sqrt{\frac{2 * 8.475}{5}}$$

$$= 3.90$$

Source $A$: = $\bar{Y}_1$ = 94.2; Source $B$: $\bar{Y}_2$ = 85; Source $C$: $\bar{Y}_3$ = 87.8 ; Source $D$: $\bar{Y}_4$ = 78

Comparison of all possible pairs of means is done as follows. We consider the absolute difference in the two means for comparison.

$\bar{Y}_1$ vs $\bar{Y}_2$ : 94.2 – 85.0 = – 9.2 > 3.9 Significant
$\bar{Y}_1$ vs $\bar{Y}_3$ : 94.2 – 87.8 = – 6.4 > 3.9 Significant
$\bar{Y}_1$ vs $\bar{Y}_4$ : 94.2 – 78.0 = – 16.2 > 3.9 Significant
$\bar{Y}_2$ vs $\bar{Y}_3$ : 85.0 – 87.8 = – 2.8 < 3.9 Not significant
$\bar{Y}_2$ vs $\bar{Y}_4$ : 85.0 – 78.0 = – 7 > 3.9 Significant
$\bar{Y}_3$ vs $\bar{Y}_4$ : 87.8 – 78.0 = – 9.8 > 3.9 Significant

This test also gives the same result.

$\underline{A}$ $\underline{BC}$ $\underline{D}$

***Turkey's test***

$$T_\alpha = q_\alpha(a, f)\sqrt{\frac{MS_e}{n}} \quad q_{0.05}(4,16) = 4.05$$

$$= 4.05\sqrt{\frac{8.475}{5}} = 5.27$$

Source $A$: $\bar{Y}_1$ = 94.2; Source $B$: $\bar{Y}_2$ = 85; Source $C$: $\bar{Y}_3$ = 87.8; Source $D$: $\bar{Y}_4$ = 78

Comparison is similar to Fisher's LSD test.

$\bar{Y}_1$ vs $\bar{Y}_2$ : 94.2 – 85.0 = – 9.2 > 5.27 Significant
$\bar{Y}_1$ vs $\bar{Y}_3$ : 94.2 – 87.8 = – 6.4 > 5.27 Significant
$\bar{Y}_1$ vs $\bar{Y}_4$ : 94.2 – 78.0 = – 16.2 > 5.27 Significant
$\bar{Y}_2$ vs $\bar{Y}_3$ : 85.0 – 87.8 = – 2.8 < 5.27 Not significant
$\bar{Y}_2$ vs $\bar{Y}_4$ : 85.0 – 78.0 = – 7 > 5.27 Significant
$\bar{Y}_3$ vs $\bar{Y}_4$ : 87.8 – 78.0 = – 9.8 > 5.27 Significant

This test also produces the same result.

$\underline{A}$ $\underline{BC}$ $\underline{D}$

We can also construct a confidence interval in this procedure using Eq. (3.29)

$$\bar{Y}_{i.} - t_{\alpha/2,N-\alpha}\sqrt{\frac{MS_E}{n}} \le \mu_2 \le \bar{Y}_{i.} + t_{\alpha/2,N-\alpha}\sqrt{\frac{MS_E}{n}} \tag{3.29}$$

### 3.2.9 Multiple Comparisons of Means Using Contrasts

A contrast is a linear combination of treatment totals and the sum of its coefficients should be equal to zero. Suppose $T_{i.}$ is the $i$th treatment total, ($i$ = 1, 2, …, $a$).

We can form a contrast, say $C_1 = T_{1.} - T_{2.}$ (3.30)

Note that in Eq. (3.29) the sum of coefficients of the treatments is zero; (+1) + (–1) = 0. Testing $C_1$ is equivalent to test the difference in two means. So, a $t$-test can be used to test the contrasts.

$$t_0 = \frac{\sum_{i=1}^{a} c_i T_{i.}}{\sqrt{nMS_e \sum_{i=1}^{a} c_i^2}} \tag{3.31}$$

where, $c_i$ is the coefficient of the $i$th treatment.

If $|t_0| > t_{\alpha/2,N-a}$, reject $H_0$

Alternatively, an $F$-test can be used.

$$F_0 = \frac{MS_C}{MS_e} \quad \text{where, } MS_C \text{ is the contrast mean square.}$$

$$= \frac{SS_C/1}{MS_e} \quad \text{(each contrast will have one degree of freedom)} \tag{3.32}$$

where,

$$SS_C = \frac{\left(\sum_{i=1}^{a} c_i T_{i.}\right)^2}{n\sum_{i=1}^{a} c_i^2} = \frac{C_i^2}{n\sum_{i=1}^{a} c_i^2} \tag{3.33}$$

where $C_i$ is the $i$th contrast totals.

Reject $H_0$, if $F_0 > F_{\alpha,1,N-a}$

### *Orthogonal contrasts*

Two contrasts with coefficients $c_i$ and $d_i$ are orthogonal if $\sum_{i=1}^{a} c_i d_i = 0$. (3.34)

In any contrast if one of the treatments is absent, its coefficient is considered as zero. For example,

$$C_1 = T_{1.} - T_{2.}$$

$$C_2 = T_{3.} - T_{4.}$$

Sum of product of the coefficients of these two contrasts $C_1$ and $C_2$ is 0

$$(1) * (0) + (-1) * (0) + (1) * (0) + (-1) * (0) = 0.$$

Thus, $C_1$ and $C_2$ are said to be orthogonal. For $a$ treatments, a set of $a$ – 1 orthogonal contrasts partition the treatment sum of squares into one degree of freedom $a$ – 1 independent components. That is, the number of contrasts should be equal to ($a$ – 1). All these contrasts should be orthogonal. Generally, the contrasts are formed prior to experimentation based on what the experimenter wants to test. This will avoid the bias if any that may result due to examination of data (if formed after experimentation).

***Testing contrasts:*** In Illustration 3.2, we have four treatments or levels ($A$, $B$, $C$ and $D$). Hence, we can form three orthogonal contrasts. Let these contrasts be,

$$C_1 = T_{1.} - T_{2.}$$
$$C_2 = T_{1.} + T_2 - T_{3.} - T_{4.}$$
$$C_3 = T_{3.} - T_{4.}$$

Note that any pair of these contrasts is orthogonal. By substituting the values of treatment totals (Table 3.10),

$$C_1 = 471 - 425 = 46$$
$$C_2 = 471 + 425 - 439 - 390 = 67$$
$$C_3 = 439 - 390 = 49$$

The sum of squares of these contrast are

$$SS_{C_1} = \frac{(C_1)^2}{n\sum c_i^2} = \frac{(46)^2}{5(2)} = 211.60$$

$$SS_{C_2} = \frac{(67)^2}{5(4)} = 224.45$$

$$SS_{C_3} = \frac{(49)^2}{5(2)} = 240.10$$

$$\text{Total} = 676.15$$

Thus, the sum of the contrast SS is equal to $SS_T$. That is, the treatment sum of squares is partitioned into three ($a$ – 1) single degree of freedom of contrast sum of squares. These are summarized in Table 3.12.

**TABLE 3.12** ANOVA for testing contrasts (Illustration 3.2)

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean squares* | $F_0$ |
|---|---|---|---|---|
| Between treatments | 676.15 | 3 | 225.38 | 26.59 |
| $C_1$ | 211.60 | 1 | 211.60 | 24.96 |
| $C_2$ | 224.45 | 1 | 224.45 | 26.48 |
| $C_3$ | 240.10 | 1 | 240.10 | 28.33 |
| Error | 135.60 | 16 | 8.475 | |
| Total | 811.75 | 19 | | |

$F_{5\%,1,16} = 4.49$, At 5% significance level, all the three contrasts are significant.

Thus, the difference in strength between *A* and *B*, and *C* and *D* is significant and also the difference between the sum of means of *A* and *B* and *C* and *D* is significant.

## 3.3 RANDOMIZED COMPLETE BLOCK DESIGN

In any experiment variability due to a nuisance factor can affect the results. A nuisance factor or noise factor affects response. A nuisance factor can be treated in experiments as shown in Figure 3.6.

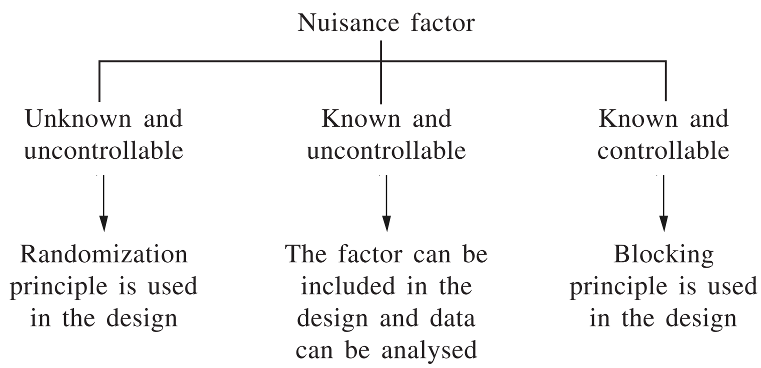



**FIGURE 3.6** A nuisance factor.

If the nuisance factor is not addressed properly in design, the error variance would be large and sometimes we may not be able to attribute whether the variation is really due to treatment. As shown in Figure 3.6, when the noise factor is known and controllable, we use randomized block design. In this design we control the variation due to one source of noise. This is explained by an example. Suppose there are four different types of drill bits used to drill a hole. We want to determine whether these four drill bits produce the same surface finish or not. If the experimenter decides to have four observations for each drill bit, he requires 16 test samples. If he assigns the samples randomly to the four drill bits, it will be a completely randomized design. If these samples are homogenous (have more or less same metallurgical properties), any variation between treatments can be attributed to the drill bits. If the samples differ in metallurgical properties, it is difficult to conclude whether the surface finish is due to the drill bits or samples and the random error will contain both error and variability between the samples.

In order to separate the variability between the samples from the error, each drill bit is used once on each of the four samples. This becomes the randomized complete block design. The samples are the blocks as they form more homogenous experimental units. Complete indicates that each block contains all the treatments (drill bits). This design is widely used in practice. The blocks can be batches of material, machines, days, people, different laboratories, etc. which contribute to variability that can be controlled. In this design the blocks represent a restriction on randomization. But within the block randomization is permitted.

### 3.3.1 Statistical Analysis of the Model

Let $a$ treatments to be compared and we have $b$ blocks.

The effects model is

$$Y_{ij} = \mu + T_i + B_j + e_{ij} \begin{cases} i = 1, 2, ..., a \\ j = 1, 2, ..., b \end{cases} \tag{3.35}$$

where,

$\mu$ = overall effect
$T_i$ = effect of $i$th treatment
$B_j$ = effect of $j$th block
$e_{ij}$ = random error
$SS_{\text{Total}} = SS_T + SS_{\text{Block}} + SS_e$
df: $N - 1 = (a - 1) + (b - 1) + (a - 1)(b - 1)$

**ILLUSTRATION 3.3**

**Randomized Complete Block Deign**

Consider a surface finish testing experiment described in Section 3.3. There are four different types of drill bits and four metal specimens. Each drill bit is used once on each specimen resulting in a randomized complete block design. The data obtained are surface roughness measurements in microns and is given in Table 3.13. The order of testing the drill bits on each specimen is random.

**TABLE 3.13** Surface roughness data (Illustration 3.3)

| *Types of drill bit* | *Specimens* (blocks) | | | | $T_{i.}$ |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | |
| 1 | 9 | 10 | 8 | 12 | 39 |
| 2 | 19 | 22 | 18 | 23 | 82 |
| 3 | 28 | 30 | 23 | 22 | 103 |
| 4 | 18 | 23 | 21 | 19 | 81 |
| $T_{.j}$ | 74 | 85 | 70 | 76 | $T_{..}$ = 305 |

***Data analysis:*** The computation required for ANOVA are as follows:

$$\text{Correction factor (CF)} = \frac{(305)^2}{16} = 5814.06$$

$$SS_{\text{Total}} = \sum_{i=1}^{a} \sum_{j=1}^{b} Y_{ij}^2 - \text{CF} \tag{3.36}$$

$$= 9^2 + 9^2 + \cdots + 9^2 + 9^2 - \text{CF} = 624.94$$

$$SS_T = \frac{\sum_{i=1}^{a} T_{i.}^2}{b} - \text{CF} \tag{3.37}$$

$$= \frac{(39)^2 + (82)^2 + (103)^2 + (81)^2}{4} - \text{CF} = 539.69$$

The block sum of squares is computed from the block totals.

$$SS_B = \frac{\sum_{j=1}^{b} T_{.j}^2}{a} - \text{CF} \tag{3.38}$$

$$= \frac{(74)^2 + (85)^2 + (70)^2 + (76)^2}{4} - \text{CF} = 30.19$$

$$SS_e = SS_{\text{Total}} - SS_T - SS_{\text{Block}} = 55.06$$

These computations are summarized in Table 3.14

**TABLE 3.14** ANOVA for the effect of types of drill bits (Illustration 3.3)

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean squares* | $F_0$ |
|---|---|---|---|---|
| Types of drill bits (Treatments) | 539.69 | 3 | 179.89 | 29.39 |
| Specimens (Blocks) | 30.19 | 3 | 10.06 | 1.64 |
| Error | 55.06 | 9 | 6.12 | |
| Total | 624.94 | 15 | | |

$F_{0.05,3,9} = 3.86$

*Conclusion:* Since $F_{0.05,3,9}$ is less than $F_0$, the null hypothesis is rejected.

That is, different types of drill bits produce different values of surface roughness.

Also, here the treatments are fixed and we can use any one multiple comparison method for comparing all pairs of treatment means. Note that, the number of replications in each treatment in this case is the number of blocks (b). The analysis of residuals can be performed for model adequacy checking as done in the case of completely randomized design. The residuals are computed as follows:

$$R_{ij} = Y_{ij} - \hat{Y}_{ij}$$

where, $\hat{Y}_{ij} = \overline{Y}_{i.} + \overline{Y}_{j.} - \overline{Y}_{..}$

So, $$R_{ij} = Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{j.} + \overline{Y}_{..} \tag{3.39}$$

### 3.3.2 Estimating Missing Values in Randomized Block Design

Sometimes we may miss an observation in one of the blocks. This may happen due to a reason beyond the control of the experimenter. When the experimental units are animals or plants, an animal/plant may die or an experimental unit may get damaged etc. In such cases, we will miss an observation. A missing observation introduces a new problem into the analysis because treatments are not orthogonal to blocks. That is, every treatment does not occur in every block. Under these circumstances we can estimate the missing observation approximately and complete the analysis, but the error degrees of freedom are reduced by one. The missing observation is estimated such that the sum of squares of errors is minimized. The table below shows the coded data for the surface finish testing experiment and suppose $X$ is the missing observation. Coded data are obtained by subtracting 20 from each observation and given in Table 3.15.

**TABLE 3.15** Coded data for Illustration 3.3 with one missing observation

| *Types of drill bit* | *Specimens* (blocks) | | | | $T_{i.}$ |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | |
| 1 | –11 | –10 | –12 | –8 | –41 |
| 2 | –1 | 2 | –2 | 3 | 2 |
| 3 | 8 | $X$ | 3 | 2 | 13 |
| 4 | –2 | 3 | 1 | –1 | 1 |
| $T_{.j}$ | –6 | –5 | –10 | –4 | –25 |

The missing observation $X$ is estimated such that the $SS_e$ are minimized.

$$SS_E = \left(\sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij}^2 - \text{CF}\right) - \frac{1}{b}\left(\sum_{i=1}^{a} Y_{i.}^2 - \text{CF}\right) - \frac{1}{a}\left(\sum_{j=1}^{b} Y_{.j}^2 - \text{CF}\right) \tag{3.40}$$

$$= \sum\sum Y_{ij}^2 - \frac{1}{b}\sum Y_{i.}^2 - \frac{1}{a}\sum Y_{.j}^2 + \frac{1}{ab} Y_{..}^2$$

$$= X^2 - \frac{1}{b}(Y'_{i.} + X)^2 - \frac{1}{a}(Y'_{.j} + X)^2 + \frac{1}{ab}(Y'_{..} + X)^2 + C \tag{3.41}$$

where $C$ includes all other terms not involving $X$. and $Y'$ represent the total without the missing observation that is,

$$Y'_{3.} = 13,\ Y'_{.2} = -5 \text{ and } Y'_{..} = -25$$

$X$ can be attained from $\dfrac{dSS_e}{dX} = 0$

$$X = \frac{aY'_{i.} + bY'_{.j} - Y'_{..}}{(a-1)(b-1)} \tag{3.42}$$

$$= \frac{4(13) + 4(-5) - (-25)}{(4 - 1)(4 - 1)} = 6.33$$

That is, in terms of original data $X = 26.33$.

The ANOVA with the estimated value for the missing observation is given in Table 3.16.

**TABLE 3.16** ANOVA with one missing observation

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean squares* | $F_0$ |
|---|---|---|---|---|
| Types of drill bits | 493.13 | 3 | 164.37 | 27.67 |
| Specimens (Blocks) | 16.66 | 3 | 5.55 | 0.93 |
| Error | 47.5 | 8 | 5.94 | |
| Total | 557.29 | 14 | | |

$F_{0.05,3,8} = 3.86$

At 5% level of significance, the treatment effect is significant.

Note that the result does not change.

## 3.4 BALANCED INCOMPLETE BLOCK DESIGN (BIBD)

If every treatment is not present in every block, it is called *randomized incomplete block design*. When all treatment comparisons are equally important, the treatment combination in each block should be selected in a balanced manner, that is, any pair of treatments occurs together the same number of times as any other pair. This type of design is called a balanced incomplete block design (BIBD). Incomplete block designs are used when there is a constraint on the resources required to conduct experiments such as the availability of experimental units or facilities etc. With reference to the surface finish testing experiment (Illustration 3.3), suppose the size of the specimen is just enough to test three drill bits only, we go for balanced incomplete block design.

Suppose in a randomized block design (day as block), four experiments are to be conducted in each block for each treatment. If only 3 experiments are possible in each day, we go for BIBD. Similarly if a batch of raw material (block) is just sufficient to conduct only three treatments out of four, we use BIBD. Tables are available for selecting BIBD designs for use (Cochran and Cox 2000).

### ILLUSTRATION 3.4

**Balanced Incomplete Block Design (BIBD)**

A company manufactures detonators using different types of compounds. The ignition time of the detonators is an important performance characteristic. They wish to study the effect of type of compound on the ignition time. The procedure is to prepare the detonators using a particular compound, test it and measure the ignition time. Currently they want to experiment with four different types of compounds. Because variation in the detonators may affect the performance of the compounds, detonators are considered as blocks. Further, each compound is just enough to prepare three detonators. Therefore, a randomized incomplete block design must be used. The design and data are given in Table 3.17. The order of the experimentation in each block is random.

**TABLE 3.17** Data for BIBD (Illustration 3.4)

| *Treatment* (*Compound*) | *Detonators* (blocks) 1 | 2 | 3 | 4 | $T_{i.}$ |
|---|---|---|---|---|---|
| 1 | 23 | 23 | 27 | — | 73 |
| 2 | 19 | 21 | — | 30 | 70 |
| 3 | 25 | — | 33 | 36 | 94 |
| 4 | — | 31 | 42 | 41 | 114 |
| $T_{.j}$ | 67 | 75 | 102 | 107 | $T_{..}$ = 351 |

### 3.4.1 Statistical Analysis of the Model

The statistical model is

$$Y_{ij} = \mu + T_i + B_j + e_{ij} \quad \begin{cases} i = 1, 2, ..., a \\ j = 1, 2, ..., b \end{cases} \tag{3.43}$$

$Y_{ij}$ is the $i$th observation in the $j$th block
$\mu$ = overall effect
$T_i$ = effect of $i$th treatment
$B_j$ = effect of $j$th block
$e_{ij}$ = NID (0, $\sigma^2$) random error.

For the statistical model considered, we have

$a$ = number of treatments = 4
$b$ = number of blocks = 4
$r$ = number of replications (observations) in each treatment = 3
$K$ = number of treatments in each block = 3

$$SS_{\text{Total}} = SS_{T(\text{adjusted})} + SS_{\text{Block}} + SS_e \tag{3.44}$$

$$SS_{\text{Total}} = \sum_i \sum_j Y_{ij}^2 - \text{CF} \left( \text{CF} = \frac{T_{..}^2}{N} \right) \tag{3.45}$$

The treatment sum of squares is adjusted to separate the treatment and block effects because each treatment is represented in a different set of $r$ blocks.

$$SS_{T(\text{adjusted})} = \frac{K \sum_{i=1}^{a} Q_i^2}{\lambda a} \tag{3.46}$$

where,

$Q_i$ = adjusted total of the $i$th treatment.

$$Q_i = Y_{i.} - \frac{1}{K}\sum_{j=1}^{b} n_{ij}Y_{.j}, \quad i = 1, 2, \ldots, a \tag{3.47}$$

$n_{ij} = 1$, if treatment appears in block $j$, otherwise $n_{ij} = 0$.
The sum of adjusted totals will be zero and has $a - 1$ degrees of freedom.

$$\lambda = \frac{r(K-1)}{a-1} \tag{3.48}$$

$$SS_{\text{Block}} = \frac{1}{K}\sum_{j=1}^{b} T_{.j}^2 - \text{CF with } b - 1 \text{ degrees of freedom.} \tag{3.49}$$

For Illustration 3.4,

$a = 4$, $b = 4$, $r = 3$, $K = 3$ and $\lambda = 2$

$$\text{CF} = \frac{(351)^2}{12} = 10266.75$$

$$SS_{\text{Total}} = 10908 - 10266.75 = 638.25$$

$$SS_{\text{Block}} = \frac{1}{3}(67^2 + 75^2 + 102^2 + 107^2) - \text{CF} = 388.92$$

Adjusted treatment totals:

$$Q_1 = T_{1.} - \frac{1}{3}\,[T_{.1} + T_{.2} + T_{.3} + 0(T_{.4})]$$

$$= 73 - \frac{1}{3}(67 + 75 + 102 + 0) = -\frac{25}{3}$$

$$Q_2 = 70 - \frac{1}{3}(67 + 75 + 0 + 107) = -\frac{39}{3}$$

$$Q_3 = 94 - \frac{1}{3}(67 + 0 + 102 + 107) = \frac{6}{3}$$

$$Q_4 = 114 - \frac{1}{3}(0 + 75 + 102 + 107) = \frac{58}{3}$$

$$SS_{T(\text{adjusted})} = \frac{K\sum_{i=1}^{a} Q_i^2}{\lambda a} \quad \text{(Eq. 3.46)}$$

$$= \frac{3\left[\left(-\frac{25}{3}\right)^2 + \left(-\frac{39}{3}\right)^2 + \left(\frac{6}{3}\right)^2 + \left(\frac{58}{3}\right)^2\right]}{2(4)} = 231.08$$

The computations are summarized in Table 3.18.

**TABLE 3.18** ANOVA for the Illustration 3.4

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean squares* | $F_0$ |
|---|---|---|---|---|
| Treatments (adjusted for blocks) | 231.08 | 3 | 77.03 | 21.10 |
| Blocks | 388.92 | 3 | 129.64 | — |
| Error | 18.25 | 5 | 3.65 | |
| Total | 638.25 | 11 | | |

Since $F_{0.05,3,9} = 5.41 < F_0$, we conclude that the compounds used has a significant effect on the ignition time.

Any multiple comparison method may be used to compare all the pairs of adjusted treatment effects.

$$\hat{T}_i = \frac{KQ_i}{\lambda a} \tag{3.50}$$

Standard error

$$(S_e) = \sqrt{\frac{K\,(MS_e)}{\lambda a}} \tag{3.51}$$

## 3.5 LATIN SQUARE DESIGN

In Randomized complete block design, we tried to control/eliminate one source of variability due to a nuisance factor. In Latin square design two sources of variability is eliminated through blocking in two directions. This design can best be explained by an example. A space research centre is trying to develop solid propellant for use in their rockets. At present they are experimenting with four different formulations. Each formulation is prepared from a batch of raw material that is just enough for testing four formulations. These formulations are prepared by different operators who differ in their skill and experience level. Thus, there are two sources of variation, one is the batch of material and the second one is the operators. Hence, the design consists of testing each formulation only once with each batch of material and each formulation to be prepared only once by each operator. Thus, the blocking principle is used to block the batch of material as well as the operators. This imposes restriction on randomization in both the directions (column wise and row wise). In this design the treatments are denoted by the Latin letters $A$, $B$, $C$, …, etc., and hence it is called Latin square design. In this design, each letter appears only once in each row and only once in each column.

$A$ 5 × 5 Latin square design is shown in Figure 3.7.

In general a Latin square of order $P$ is a $P \times P$ square of $P$ Latin letters such that each Latin letter appears only once in a row and only once in each column. The levels of the two blocking factors are assigned randomly to the rows and columns and the treatments of the experimental factor are randomly assigned to the Latin letters. Selection of Latin square designs is discussed in Fisher and Yates (1953).

| | | | | |
|---|---|---|---|---|
| *A* | *B* | *C* | *D* | *E* |
| *C* | *D* | *E* | *A* | *B* |
| *D* | *E* | *A* | *B* | *C* |
| *E* | *A* | *B* | *C* | *D* |
| *B* | *C* | *D* | *E* | *A* |

**FIGURE 3.7** 5 × 5 Latin square design.

### 3.5.1 The Statistical Model

$$Y_{ijk} = \mu + A_i + T_k + B_j + e_{ijk} \begin{cases} i = 1, 2, ..., P \\ j = 1, 2, ..., P \\ k = 1, 2, ..., P \end{cases} \tag{3.52}$$

$Y_{ijk}$ = observation in the *i*th row and *j*th column for the *k*th treatment
$\mu$ = overall mean
$A_i$ = row effect
$T_k$ = effect of *k*th treatment
$B_j$ = column effect
$e_{ijk}$ = NID(0, $\sigma^2$) random error.

The total sum of squares is partitioned as

$$SS_{\text{Total}} = SS_{\text{row}} + SS_{\text{col}} + SS_T + SS_e \tag{3.53}$$

$$\text{df: } P^2 - 1 = (P - 1) + (P - 1) + (P - 1) + (P - 2)\,(P - 1) \tag{3.54}$$

The appropriate statistic for testing no difference in treatment means is

$$F_0 = \frac{MS_T}{MS_e} \tag{3.55}$$

The standard format for ANOVA shall be used.

**ILLUSTRATION 3.5**

**Latin Square Design**

Suppose for the solid propellant experiment described in Section 3.5, the thrust force developed (coded data) from each formulation (*A*, *B*, *C* and *D*) is as follows (Table 3.19)

The treatment totals are obtained by adding all *A*s, all *B*s, etc.

$$A = 75,\ B = 43,\ C = 58,\ D = 22$$

**TABLE 3.19** Data for the Illustration 3.5 (Latin Square Design)

| *Batch of raw material* | *Operators* 1 | 2 | 3 | 4 | *Row total* |
|---|---|---|---|---|---|
| 1 | $A$ = 14 | $D$ = 6 | $C$ = 10 | $B$ = 8 | 38 |
| 2 | $D$ = 4 | $C$ = 14 | $B$ = 10 | $A$ = 19 | 47 |
| 3 | $C$ = 18 | $B$ = 11 | $A$ = 22 | $D$ = 7 | 58 |
| 4 | $B$ = 14 | $A$ = 20 | $D$ = 5 | $C$ = 16 | 55 |
| Column total | 50 | 51 | 47 | 50 | $T_{...}$ = 198 |

The correction factor (CF) = $\dfrac{(198)^2}{16}$ = 2450.25

$$SS_T = \frac{(75)^2 + (43)^2 + (58)^2 + (22)^2}{4} - \text{CF}$$

$$= 2830.5 - 2450.25 = 380.25$$

using row totals,

$$SS_{\text{row}} = 2510.5 - \text{CF} = 60.25$$

from column totals,

$$SS_{\text{col}} = 2452.5 - \text{CF} = 2.25$$

$$SS_{\text{Total}} = 2924.0 - \text{CF} = 473.75$$

$$SS_e = 473.75 - 442.75 = 31.0$$

These computations are summarized in Table 3.20.

**TABLE 3.20** ANOVA for Latin Square Design (Illustration 3.5)

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean squares* | $F_0$ |
|---|---|---|---|---|
| Formulation (Treatment) | 380.25 | 3 | 126.75 | 24.52 |
| Batch of material | 60.25 | 3 | 20.08 | — |
| Operators | 2.25 | 3 | 0.75 | — |
| Error | 31.00 | 6 | 5.17 | — |
| Total | 473.75 | 15 | | |

$F_{0.05,3,6}$ = 4.46

Since $F_0 > F_{0.05,3,6}$ = 4.46, the type of formulation has significant effect on the thrust force developed.

## 3.6 GRAECO-LATIN SQUARE DESIGN

In Latin square design, we try to control variability from two nuisance variables by blocking through rows and columns. In any single-factor experiment if three nuisance variables are present, we use Graeco-Latin Square Design. In this design, in addition to rows and columns, we use Greek letters ($\alpha$, $\beta$, $\gamma$, $\delta$, etc.) to represent the third nuisance variable. By superimposing a $P \times P$ Latin square on a second $P \times P$ Latin square with the Greek letters such that each Greek letter appears once and only once with each Latin letter, we obtain a Graeco-Latin square design. This design allows investigating four factors (rows, columns, and Latin and Greek letters) each at $P$-levels in only $P^2$ runs. Graeco-Latin squares exist for all $P \geq 3$, except $P = 6$. An example of a $4 \times 4$ Graeco-Latin square design is given in Table 3.21.

**TABLE 3.21** $A$ 4 × 4 Graeco-Latin square design

| *Rows* | *Columns* | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| 1 | $A\ \alpha$ | $B\ \beta$ | $C\ \gamma$ | $D\ \delta$ |
| 2 | $B\ \delta$ | $A\ \gamma$ | $D\ \beta$ | $C\ \alpha$ |
| 3 | $C\ \beta$ | $D\ \alpha$ | $A\ \delta$ | $B\ \gamma$ |
| 4 | $D\ \gamma$ | $C\ \delta$ | $B\ \alpha$ | $A\ \beta$ |

### 3.6.1 The Statistical model

$$Y_{ijkl} = \mu + A_i + B_j + C_k + D_l + e_{ijkl} \quad \begin{cases} i = 1, 2, ..., P \\ j = 1, 2, ..., P \\ k = 1, 2, ..., P \\ l = 1, 2, ..., P \end{cases} \tag{3.56}$$

where,

$\mu$ = overall effect
$A_i$ = effect of $i$th row
$B_j$ = effect of $j$th column
$C_k$ = effect of Latin letter treatment $k$
$D_l$ = effect of Greek letter treatment $l$
$e_{ijkl}$ = NID(0, $\sigma^2$) random error

The analysis of variance is similar to Latin square design. The sum of squares due to the Greek letter factor is computed using Greek letter totals.

## PROBLEMS

**3.1** The effect of temperature on the seal strength of a certain packaging material is being investigated. The temperature is varied at five different fixed levels and observations are given in Table 3.22.

**TABLE 3.22** Data for Problem 3.1

| *Temperature* (°C) | *Seal strength* ($N/mm^2$) | | | |
|---|---|---|---|---|
| 105 | 5.0 | 5.5 | 4.5 | 4.9 |
| 110 | 6.9 | 7.0 | 7.1 | 7.2 |
| 115 | 10.0 | 10.1 | 10.5 | 10.4 |
| 120 | 12.2 | 12.0 | 12.1 | 12.1 |
| 125 | 9.4 | 9.3 | 9.6 | 9.5 |

(a) Test the hypothesis that temperature affects seal strength. Use $\alpha = 0.05$.
(b) Suggest the best temperature to be used for maximizing the seal strength (use Duncan's multiple range test).
(c) Construct a normal probability plot of residuals and comment.
(d) Construct a 95% confidence interval for the mean seal strength corresponding to the temperature 120°C.

**3.2** Re-do part (b) of Problem 3.1 using Newman–Keuls test. What conclusions will you draw?

**3.3** A study has been conducted to test the effect of type of fixture (used for mounting an automotive break cylinder) on the measurements obtained during the cylinder testing. Five different types of fixtures were used. The data obtained are given in Table 3.23. The data are the stroke length in mm. Five observations were obtained with each fixture. Test the hypothesis that the type of fixture affects the measurements. Use $\alpha = 0.05$.

**TABLE 3.23** Data for Problem 3.3

| *Type of fixture* | | | | |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |
| 1.24 | 1.30 | 1.21 | 1.18 | 1.12 |
| 1.24 | 1.24 | 1.21 | 1.12 | 1.18 |
| 1.09 | 1.31 | 1.33 | 1.25 | 1.22 |
| 1.18 | 1.13 | 1.10 | 1.31 | 1.13 |
| 1.25 | 1.20 | 1.22 | 1.07 | 1.14 |

**3.4** An experiment was conducted to determine the effect of type of fertilizer on the growth of a certain plant. Three types of fertilizers were tried. The data for growth of the plants (cm) measured after 3 months from the date of seedling is given in Table 3.24.

**TABLE 3.24** Data for Problem 3.4

| *Type of fertilizer* | | |
|---|---|---|
| 1 | 2 | 3 |
| 12.4 | 16.1 | 31.5 |
| 13.1 | 16.4 | 30.4 |
| 11.1 | 17.2 | 31.6 |
| 12.2 | 16.6 | 32.1 |
| 11.3 | 17.1 | 31.8 |
| 12.1 | 16.8 | 30.9 |

(a) Test whether the type of fertilizer has an effect on the growth of the plants. Use $\alpha = 0.05$.
(b) Find the $p$-value for the $F$-statistic in part (a).
(c) Use Tukey's test to compare pairs of treatment means. Use $\alpha = 0.01$.

**3.5** An experiment was run to investigate the influence of DC bias voltage on the amount of silicon dioxide etched from a wafer in a plasma etch process. Three different levels of DC bias were studied and four replicates were run in a random order resulting the data (Table 3.25).

**TABLE 3.25** Data for Problem 3.5

| *DC bias* (volts) | *Amount etched* | | | |
|---|---|---|---|---|
| 398 | 283.5 | 236.0 | 231.5 | 228.0 |
| 485 | 329.0 | 330.0 | 336.0 | 384.5 |
| 572 | 474.0 | 477.5 | 470.0 | 474.5 |

Analyse the data and draw conclusions. Use $\alpha = 0.05$.

**3.6** A study was conducted to assess the effect of type of brand on the life of shoes. Four brands of shoes were studied and the following data were obtained (Table 3.26).

**TABLE 3.26** Data for Problem 3.6

| *Shoe life in months* | | | |
|---|---|---|---|
| *Brand* 1 | *Brand* 2 | *Brand* 3 | *Brand* 4 |
| 12 | 14 | 20 | 18 |
| 14 | 17 | 21 | 16 |
| 11 | 13 | 26 | 21 |
| 13 | 19 | 24 | 20 |
| 10 | 14 | 22 | 18 |

(a) Are the lives of these brands of shoes different? Use $\alpha = 0.05$.
(b) Analyse the residuals from this study.
(c) Construct a 95% confidence interval on the mean life of Brand 3.
(d) Which brand of shoes would you select for use? Justify your decision.

**3.7** Four different fuels are being evaluated based on the emission rate. For this purpose four IC engines have been used in the study. The experimenter has used a completely randomized block design and obtained data are given in Table 3.27. Analyse the data and draw appropriate conclusions. Use $\alpha = 0.05$.

**TABLE 3.27** Data for Problem 3.7

| *Type of fuel* | *Engines (Block)* | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| $F_1$ | 0.25 | 0.21 | 0.15 | 0.52 |
| $F_2$ | 0.45 | 0.36 | 0.81 | 0.66 |
| $F_3$ | 0.50 | 0.76 | 0.54 | 0.61 |
| $F_4$ | 0.28 | 0.14 | 0.11 | 0.30 |

**3.8** Four different printing processes are being compared to study the density that can be reproduced. Density readings are taken at different dot percentages. As the dot percentage is a source of variability, a completely randomized block design has been used and the data obtained are given in Table 3.28. Analyse the data and draw the conclusions. Use $\alpha = 0.05$.

**TABLE 3.28** Data for Problem 3.8

| *Type of process* | *Dot percentages (Block)* | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Offset | 0.90 | 0.91 | 0.91 | 0.92 |
| Inkjet | 1.31 | 1.32 | 1.33 | 1.34 |
| Dyesub | 1.49 | 1.54 | 1.67 | 1.69 |
| Thermalwax | 1.07 | 1.19 | 1.38 | 1.39 |

**3.9** A software engineer wants to determine whether four different types of Relational Algebraic-joint operation produce different execution time when tested on 5 different queries. The data given in Table 3.29 is the execution time in seconds. Analyse the data and give your conclusions. Use $\alpha = 0.05$.

**TABLE 3.29** Data for Problem 3.9

| *Type* | *Queries (Block)* | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| 1 | 1.3 | 1.6 | 0.5 | 1.2 | 1.1 |
| 2 | 2.2 | 2.4 | 0.4 | 2.0 | 1.8 |
| 3 | 1.8 | 1.7 | 0.6 | 1.5 | 1.3 |
| 4 | 3.9 | 4.4 | 2.0 | 4.1 | 3.4 |

**3.10** An oil company wants to test the effect of four different blends of gasoline (*A*, *B*, *C*, *D*) on fuel efficiency. The company has used four cars for testing the four types of fuel. To control the variability due to the cars and the drivers, Latin square design has been used and the data collected are given in Table 3.30. Analyse the data from the experiment and draw conclusions. Use $\alpha = 0.05$.

**TABLE 3.30** Data for Problem 3.10

| *Driver* | *Cars* | | | |
|---|---|---|---|---|
| | I | II | III | IV |
| 1 | $D = 15.5$ | $B = 33.9$ | $C = 13.2$ | $A = 29.1$ |
| 2 | $B = 16.3$ | $C = 26.6$ | $A = 19.4$ | $D = 22.8$ |
| 3 | $C = 10.8$ | $A = 31.1$ | $D = 17.1$ | $B = 30.3$ |
| 4 | $A = 14.7$ | $D = 34.0$ | $B = 19.7$ | $C = 21.6$ |

**3.11** An Industrial engineer is trying to assess the effect of four different types of fixtures (*A*, *B*, *C*, *D*) on the assembly time of a product. Four operators were selected for the study and each one of them assembled one product on each fixture. To control operator variability and variability due to product, Latin square design was used. The assembly times in minutes are given in Table 3.31. Analyse the data from the experiment and draw conclusions. Use $\alpha = 0.05$.

**TABLE 3.31** Data for Problem 3.11

| *Operator* | *Product* | | | |
|---|---|---|---|---|
| | I | II | III | IV |
| 1 | $A = 2.2$ | $B = 1.6$ | $C = 3.6$ | $D = 4.2$ |
| 2 | $B = 1.8$ | $A = 3.4$ | $D = 4.4$ | $C = 2.8$ |
| 3 | $C = 3.2$ | $D = 4.2$ | $A = 2.6$ | $B = 1.8$ |
| 4 | $D = 4.4$ | $C = 2.7$ | $B = 1.5$ | $A = 3.9$ |

**3.12** Suppose in Problem 3.11 an additional source of variation due to the work place layout/ arrangement of parts used by each operator is introduced. Data has been collected using the following Graeco-Latin square design (Table 3.32). Analyse the data from the experiment and draw conclusions. Use $\alpha = 0.05$.

**TABLE 3.32** Data for Problem 3.12

| *Operator* | *Product* | | | |
|---|---|---|---|---|
| | I | II | III | IV |
| 1 | $\alpha A = 2.5$ | $\beta C = 4.0$ | $\gamma B = 3.4$ | $\delta D = 5.3$ |
| 2 | $\beta D = 5.6$ | $\alpha B = 3.8$ | $\delta C = 4.5$ | $\gamma A = 2.1$ |
| 3 | $\gamma C = 4.7$ | $\delta A = 2.7$ | $\alpha D = 5.9$ | $\beta B = 3.8$ |
| 4 | $\delta B = 3.2$ | $\gamma D = 5.1$ | $\beta A = 2.2$ | $\alpha C = 4.4$ |

CHAPTER 4

# Multi-factor Factorial Experiments

## 4.1 INTRODUCTION

In a single-factor experiment, only one factor is studied. And the levels of the factor are the treatments. The main purpose of these experiments is to compare the treatments in all possible pairs in order to select the best treatment or its alternatives. When the number of factors involved in the experiment is more than one, we call it a *factorial experiment*. In factorial experiments, combination of two or more levels of more than one factor is the treatment. That is, every level of one factor is combined with every level of other factors. When all the possible treatments are studied, we call it a *full factorial experiment*. If the number of factors is only two, it will be a two-factor factorial experiment.

## 4.2 TWO-FACTOR EXPERIMENTS

Suppose we want to study the effect of temperature and pressure on the reaction time of a chemical process. Further we want to investigate the temperature at two levels (70°C and 90°C) and pressure at two levels (200 MPa and 250 MPa). The two-factor factorial design will be represented as given in Table 4.1.

**TABLE 4.1** Two-factor factorial design

| *Pressure* (Mpa) | *Temperature* (°C) | |
|---|---|---|
| | 60 | 90 |
| 200 | 1 | 2 |
| 250 | 3 | 4 |

Note that there are four treatment combinations (1, 2, 3 and 4) in the two-factor design as given in Table 4.1.

In factorial experiments, we are interested in testing the main effects as well as interaction effects. Let us study these two aspects with an example. Suppose in the above two-factor experiment, the reaction time at the four treatment combinations can be graphically represented as shown in

Figures 4.1(a) and 4.1(b). Usually in a two-level factorial design, we represent the levels of the factor as low (–) and high (+).

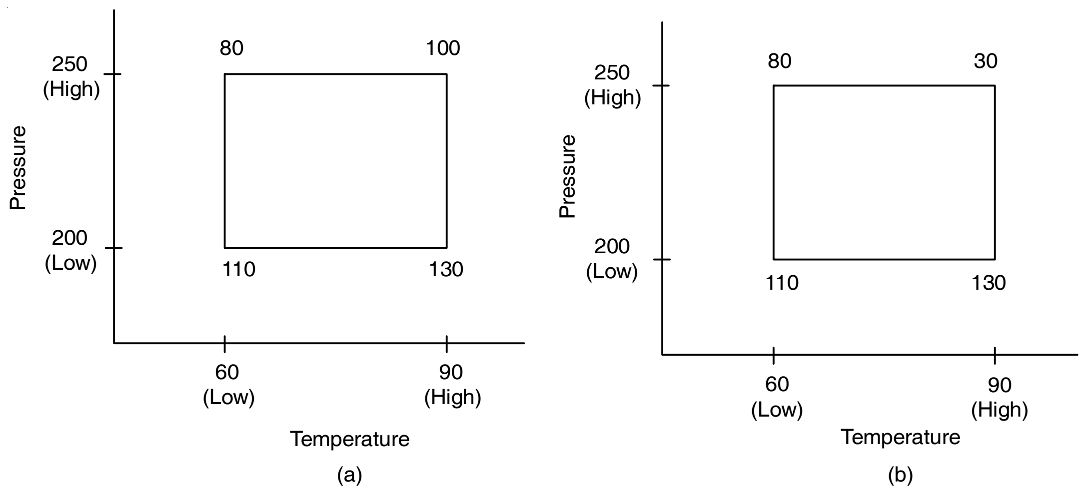



**FIGURE 4.1** Representation of response data for a two-factor experiment.

***Main effect of a factor:*** The effect of a factor is defined as the change in response due to a change in the level of the factor. This is called the main effect of a factor because it pertains to the basic factor considered in the experiment. For example in the above experiment, the effect due to temperature and the effect due to pressure are the main effects. The effect due to temperature in this experiment (Figure 4.1(a)) is the average of the effect at the high level pressure and at the low level of pressure. That is,

Effect of temperature at the high level of pressure = (100 – 80) = 20
Effect of temperature at the low level of pressure = (130 – 110) = 20
Average effect of temperature = 1/2(20 + 20) = 20

Similarly, we can find the effect of pressure.
Average effect of pressure = 1/2{(100 – 130) + (80 – 100)} = –30

***Interaction effect:*** The combined effect due to both the factors is called the *interaction effect*. From Figure 4.2(a), it is observed that the behaviour of response at the two levels of the other factor is same. That is, when temperature changes from 60°C to 90°C (Figure 4.1(a)), the response increases at the two levels of the other factor, pressure. Hence, we say that there is no interaction effect. On the other hand, from Figure 4.1(b), it is observed that, when temperature changes from 60°C to 90°C, at the low level of the factor, pressure, the response increases and at the high level the response decreases. Thus, the behaviour of response is different at two levels and we say that there is an interaction effect. Interaction effect is computed as the average difference between the factor effects explained as follows:

From Figure 4.1(b),

Temperature effect at the low level of pressure = 130 – 110 = 20
Temperature effect at the high level of pressure = 30 – 80 = –50
Interaction effect = 1/2{(–50) – (20)} = –35

With respect to the factor (pressure) also, we can compute interaction effect and it will be same.

Pressure effect at the low level of temperature = 80 – 110 = –30
Pressure effect at the high level of temperature = 30 – 130 = –100
Interaction effect = 1/2{(–100) – (–30)} = –35

The presence of interaction effect is illustrated graphically in Figure 4.2a and 4.2b. Figures 4.2a and 4.2b shown below are the plots of response data from Figures 4.1a and 4.1b respectively.

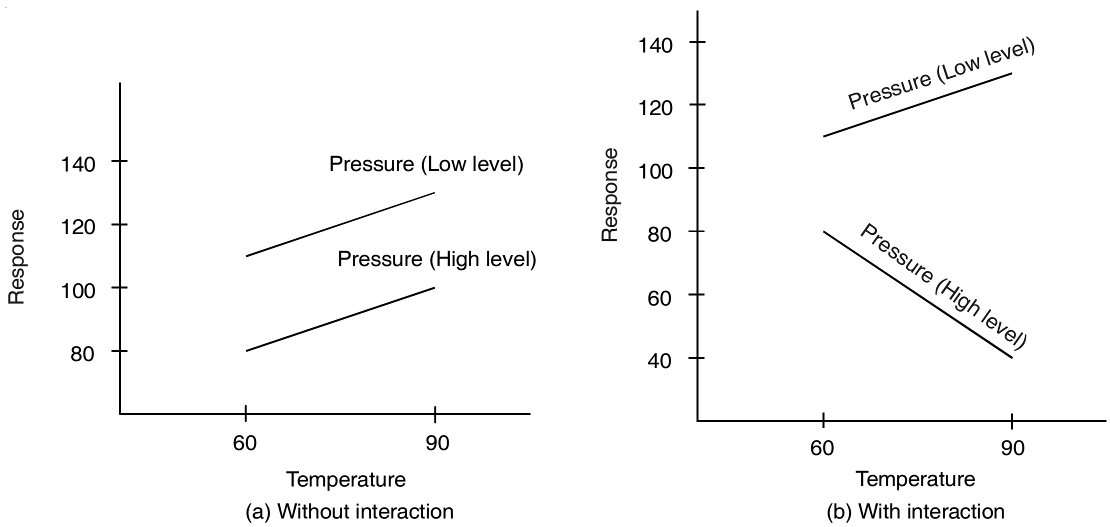



**FIGURE 4.2** Plot of response data for the two-factor experiment shown in Figure 4.1.

Note that the two response lines in Figure 4.2a are approximately parallel indicating that there is no interaction effect between temperature and pressure. Whereas in Figure 4.2b, the two response lines are not parallel indicating the presence of interaction effect. However, the presence or absence of interaction effects in factorial experiments can be ascertained through testing using ANOVA.

### 4.2.1 The Statistical Model for a Two-factor Experiment

Let the two factors be denoted by $A$ and $B$ and both are fixed.

The effects model is

$$Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + e_{ijk} \quad \begin{cases} i = 1,2, ..., a \\ j = 1,2, ..., b \\ k = 1,2, ..., n \end{cases} \tag{4.1}$$

$\mu$ = overall effect
$A_i$ = effect of $i$th level of factor $A$

$B_j$ = effect of $j$th level of factor $B$
$AB_{ij}$ = effect of interaction between $A$ and $B$
$e_{ijk}$ = random error component

In this model, we are interested in testing the main effects of $A$ and $B$ and also $AB$ interaction effect.

The hypotheses for testing are as follows:

***Effect of A:***
$H_0$: $A_1 = A_2 = \cdots = A_a = 0$ (4.2)
$H_1$: at least one $A_i \neq 0$

***Effect of B:***
$H_0$: $B_1 = B_2 = \cdots = B_b = 0$ (4.3)
$H_1$: at least one $B_j \neq 0$

***Interaction effect:***
$H_0$: $(AB)_{ij} = 0$ for all $i$, $j$ (4.4)
$H_1$: at least one $(AB)_{ij} \neq 0$

Now, we shall discuss how these effects are tested using ANOVA.

The ANOVA equation for this model is:

$$SS_{\text{Total}} = SS_A + SS_B + SS_{AB} + SS_e \tag{4.5}$$

and the degrees of freedom is:

$$abn - 1 = (a - 1) + (b - 1) + (a - 1)(b - 1) + ab(n - 1) \tag{4.6}$$

***Computation of sum of squares:***

$$\text{Correction factor (CF)} = \frac{T_{...}^2}{N} \tag{4.7}$$

where, $T_{...}$ = Grand total and $N = abn$

$$SS_{\text{Total}} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} Y_{ijk}^2 - \text{CF} \tag{4.8}$$

$$SS_A = \sum_{i=1}^{a} \frac{T_{i..}^2}{bn} - \text{CF} \tag{4.9}$$

$$SS_B = \sum_{j=1}^{b} \frac{T_{.j.}^2}{an} - \text{CF} \tag{4.10}$$

$$SS_{AB} = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{T_{ij.}^2}{n} - \text{CF} - SS_A - SS_B \tag{4.11}$$

$$SS_e = SS_{\text{Total}} - (SS_A + SS_B + SS_{AB}) \tag{4.12}$$

Table 4.2 gives the summary of ANOVA computations for a two-factor factorial fixed effect model.

**TABLE 4.2** The ANOVA for a two-factor factorial experiment

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean squares* | $F_0$ |
|---|---|---|---|---|
| $A$ | $SS_A$ | $a - 1$ | $\frac{SS_A}{a-1}$ | $\frac{MS_A}{MS_e}$ |
| $B$ | $SS_B$ | $b - 1$ | $\frac{SS_B}{b-1}$ | $\frac{MS_B}{MS_e}$ |
| $AB$ | $SS_{AB}$ | $(a - 1)(b - 1)$ | $\frac{SS_{AB}}{(a-1)(b-1)}$ | $\frac{MS_{AB}}{MS_e}$ |
| Error | $SS_e$ | $ab(n - 1)$ | $\frac{SS_e}{ab(n-1)}$ | |
| Total | $SS_{\text{Total}}$ | $abn - 1$ | | |

### 4.2.2 Estimation of Model Parameters

The parameters of the model are estimated as follows. The model for the two-factor factorial experiment is:

$$Y_{ijk} = \mu + A_i + B_j + AB_{ij} + e_{(ij)k} \tag{4.13}$$

The overall effect $\mu$ is estimated by the grand mean. The deviation of the factor mean from the grand mean is the factor effect. The various estimates are discussed as follows:

$$\hat{\mu} = \overline{Y}_{...} \tag{4.14}$$

$$\hat{A}_i = \overline{Y}_{i..} - \overline{Y}_{...} \qquad i = 1, 2, ..., a \tag{4.15}$$

$$\hat{B}_j = \overline{Y}_{.j.} - \overline{Y}_{...} \qquad j = 1, 2, ..., b \tag{4.16}$$

$$(\hat{A}B)_{ij} = \overline{Y}_{ij.} - \overline{Y}_{...} - \hat{A}_i - \hat{B}_j$$

$$= \overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...} \tag{4.17}$$

$$\hat{Y}_{ijk} = \hat{\mu} + \hat{A}_i + \hat{B}_j + (\hat{A}B)_{ij} \tag{4.18}$$

$$= \overline{Y}_{ij.} \text{ (on substitution)} \tag{4.19}$$

***Estimation of residuals*** $(R_{ijk})$**:** The residual is the difference between the observed value and the predicted value.

$$R_{ijk} = Y_{ijk} - \hat{Y}_{ijk} \tag{4.20}$$

$$= Y_{ijk} - \bar{Y}_{ij.} \tag{4.21}$$

***Analysis of residuals:*** The following graphical plots are obtained to check the adequacy of the model.

1. Plot of the residuals ($R_{ijk}$ s) on a normal probability paper; if the plot resembles a straight line, indicate that the errors are normally distributed.
2. Plot of residuals versus order of experimentation; if the residuals are randomly scattered and no pattern is observed, indicate that the errors are independent.
3. Plot of residuals versus fitted values ($R_{ijk}$ vs $\hat{Y}_{ijk}$); if the model is correct and assumptions are valid, the residuals should be structure less.
4. Plot of residuals versus the factors ($A_i$ and $B_j$); these plots indicate equality of variance of the respective factors.

## ILLUSTRATION 4.1

**Two-factor Factorial Experiment**

A chemical engineer has conducted an experiment to study the effect of temperature and pressure on the reaction time of a chemical process. The two factors are investigated at three levels each. The following data (Table 4.3) are obtained from two replications.

**TABLE 4.3** Data for Illustration 4.1

| *Pressure* (MPa) | *Temperature* (°C) 100 | | 120 | | 140 | | $T_{i..}$ |
|---|---|---|---|---|---|---|---|
| 100 | 23 | | 31 | | 36 | | |
| | 25 | 48 | 32 | 63 | 39 | 75 | 186 |
| 110 | 35 | | 34 | | 31 | | |
| | 36 | 71 | 35 | 69 | 34 | 65 | 205 |
| 120 | 28 | | 27 | | 26 | | |
| | 27 | 55 | 25 | 52 | 24 | 50 | 157 |
| $T_{.j.}$ | | 174 | | 184 | | 190 | $T_{...}$ = 548 |

***Data analysis:*** First the row totals ($T_{i..}$), column totals ($T_{.j.}$) and the cell totals (the underlined numbers) are obtained as given in Table 4.3.

***Computation of sum of squares:***

Grand total ($T_{...}$) = 548
Total number of observations ($N = abn$) = 18

$$\text{Correction factor (CF)} = \frac{T_{...}^2}{N} = \frac{(548)^2}{18} = 16683.56$$

$SS_{\text{Total}}$ is computed from the individual observations using Eq. (4.8).

$$SS_{\text{Total}} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} Y_{ijk}^2 - \text{CF}$$

$$= \{(23)^2 + (25)^2 + (31)^2 + \cdots + (26)^2 + (24^2)\} - \text{CF}$$

$$= 17094.00 - 16683.56 = 410.44$$

Sum of squares due to the factor pressure is computed using its level totals [Eq. (4.9)].

$$SS_{\text{Pressure}} = \sum_{i=1}^{a} \frac{T_{i..}^2}{bn} - \text{CF}$$

$$= \frac{(186)^2 + (205)^2 + (157)^2}{6} - 16683.56 = 194.77$$

Sum of squares due to the factor temperature is computed using its level totals [Eq. (4.10)].

$$SS_{\text{Temperature}} = \sum_{j=1}^{b} \frac{T_{.j.}^2}{an} - \text{CF}$$

$$= \frac{(174)^2 + (184)^2 + (190)^2}{6} - 16683.56 = 21.77$$

Sum of squares due to the interaction between pressure and temperature is computed using the combined totals/cell totals {Eq. (4.11)}.

$$SS_{\text{Pressure} \times \text{Temperature}} = \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{T_{ij.}^2}{n} - \text{CF} - SS_{\text{Pressure}} - SS_{\text{Temperature}}$$

$$= \frac{(48)^2 + (63)^2 + ... + (52)^2 + (50)^2}{2} - 16683.56 - 194.77 - 21.77$$

$$= 17077.00 - 16900.100 = 176.90$$

The ANOVA for the two-factor experiment (Illustration 4.1) is given in Table 4.4.

**TABLE 4.4** ANOVA for Illustration 4.1 (Two-factor experiment)

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean squares* | $F_0$ |
|---|---|---|---|---|
| Temperature | 21.77 | 2 | 10.89 | 5.73 |
| Pressure | 194.77 | 2 | 97.39 | 51.26 |
| Temperature × Pressure | 176.90 | 4 | 44.23 | 23.28 |
| Interaction Error | 17.00 | 9 | 1.90 | |
| Total | 410.44 | 17 | | |

Since $F_{0.05,2,9} = 4.26$ and $F_{0.05,4,9} = 3.63$, the main effects temperature and pressure and also their interaction effect is significant.

***Residuals:*** The residuals are computed using Eq. (4.21). Table 4.5 gives the residuals for Illustration 4.1.

**TABLE 4.5** Residuals for Illustration 4.1

| *Pressure* (MPa) | *Temperature* (°C) 100 | 120 | 140 |
|---|---|---|---|
| 100 | –1.0 | –0.5 | –1.5 |
| | 1.0 | 0.5 | 1.5 |
| 110 | 0.5 | –0.5 | –1.5 |
| | –0.5 | 0.5 | 1.5 |
| 120 | 0.5 | 1.0 | 1.0 |
| | –0.5 | –1.0 | –1.0 |

Using residuals in Table 4.5, the following graphical plots should be examined for the model adequacy as already discussed earlier. We leave this as an exercise to the reader.

1. Plot of residuals on a normal probability paper or its standardized values on an ordinary graph sheet
2. Plot of residuals vs fitted values ($\hat{Y}_{ijk}$)
3. Plot of residuals vs temperature
4. Plot of residuals vs pressure

***Comparison of means:*** When the ANOVA shows either row factor or column factor as significant, we compare the means of row or column factor to identify significant differences. On the other hand if the interaction is significant (irrespective of the status of row and column factor), the combined treatment means (cell means) are to be compared. For comparison any one pair wise comparison method can be used.

In Illustration 4.1, the interaction is also significant. Hence, the cell means have to be compared. Note that there are 36 possible pairs of means of the 9 cells.

## 4.3 THE THREE-FACTOR FACTORIAL EXPERIMENT

The concept of two-factor factorial design can be extended to any number of factors. Suppose we have factor $A$ with $a$ levels, factor $B$ with $b$ levels and factor $C$ with $c$ levels and so on, arranged in a factorial experiment. This design will have $abc \ldots n$ total number of observations, where $n$ is the number of replications. In order to obtain experimental error required to test all the main effects and interaction effects, we must have a minimum of two replications.

If all the factors are fixed, we can formulate the appropriate hypotheses and test all the effects. In a fixed effects model (all factors are fixed), the $F$-statistic is computed by dividing the mean square of all the effects by the error mean square. The number of degrees of freedom for any main effect is the number of levels of the factor minus one, and the degrees of freedom for an interaction is the product of the degrees of freedom associated with the individual effects of the interaction.

### 4.3.1 The Statistical Model for a Three-factor Experiment

Consider a three-factor factorial experiment with factors $A$, $B$ and $C$ at levels $a$, $b$ and $c$ respectively. It is customary to represent the factor by a upper case and its levels by its lower case.

$$Y_{ijkl} = \mu + A_i + B_j + (AB)_{ij} + C_k + (AC)_{ik} + (BC)_{jk} + (ABC)_{ijk} + e_{(ijk)\,l} \quad \begin{cases} i = 1, 2, ..., a \\ j = 1, 2, ..., b \\ k = 1, 2, ..., c \\ l = 1, 2, ..., n \end{cases} \tag{4.22}$$

The analysis of the model is discussed as follows:

***Computation of sum of squares:***

$$\text{Correction factor (CF)} = \frac{T_{...}^2}{abcn} \tag{4.23}$$

$$SS_{\text{Total}} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{l=1}^{n} Y_{ijkl}^2 - \text{CF} \tag{4.24}$$

The sum of squares for the main effects is computed using their respective level totals.

$$SS_A = \frac{\sum_{i=1}^{a} T_{i...}^2}{bcn} - \text{CF} \tag{4.25}$$

$$SS_B = \frac{\sum_{j=1}^{b} T_{.j..}^2}{acn} - \text{CF} \tag{4.26}$$

$$SS_C = \frac{\sum_{k=1}^{c} T_{..k.}^2}{abn} - \text{CF} \tag{4.27}$$

Using $AB$, $AC$ and $BC$ combined cell totals; the two-factor interaction sum of squares is computed.

$$SS_{AB} = \frac{\sum_{i=1}^{a} \sum_{j=1}^{b} T_{ij..}^2}{cn} - \text{CF} - SS_A - SS_B \tag{4.28}$$

$$SS_{AC} = \frac{\sum_{i=1}^{a}\sum_{k=1}^{c} T_{i.k.}^2}{bn} - \text{CF} - SS_A - SS_C \tag{4.29}$$

$$SS_{BC} = \frac{\sum_{j=1}^{b}\sum_{k=1}^{c} T_{.jk.}^2}{an} - \text{CF} - SS_B - SS_C \tag{4.30}$$

The three-factor interaction sum of squares is computed using the *ABC* combined cell totals.

$$SS_{ABC} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} T_{ijk.}^2}{n} - \text{CF} - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC} \tag{4.31}$$

The error sum of squares is obtained by subtracting the sum of squares of all main and interaction effects from the total sum of squares.

$$SS_E = SS_{\text{Total}} - SS \text{ of all effects} \tag{4.32}$$

## ILLUSTRATION 4.2

An industrial engineer has studied the effect of speed ($B$), feed ($C$) and depth of cut ($A$) on the surface finish of a machined component using a three-factor factorial design. All the three factors were studied at two-levels each. The surface roughness measurements (microns) from two replications are given in Table 4.6.

**TABLE 4.6** Surface roughness data for Illustration 4.2

| *Depth of cut* ($A$) | *Speed* ($B$) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 100 | | | | 120 | | | | |
| | *Feed* ($C$) | | | | *Feed* ($C$) | | | | |
| | 0.20 | | 0.25 | | 0.20 | | 0.25 | | $T_{i...}$ |
| 0.15 | 54 | | 41 | | 59 | | 43 | | |
| | 52 | 106 | 58 | 99 | 61 | 120 | 55 | 98 | 423 |
| 0.20 | 86 | | 62 | | 82 | | 65 | | |
| | 82 | 168 | 64 | 126 | 75 | 157 | 77 | 142 | 593 |
| $T_{.jk.}$ | | 274 | | 225 | | 277 | | 240 | |
| $T_{.j..}$ | | | | 499 | | | | 517 | 1016($T_{....}$) |

***Computation of sum of squares:***

In this example, we have $a = 2$, $b = 2$, $c = 2$ and $n = 2$

$$\text{Correction factor (CF)} = \frac{T_{....}^2}{abcn} = \frac{(1016)^2}{16} = 64516.00$$

The total sum of squares is found from the individual data using Eq. (4.24).

$$SS_{\text{Total}} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{l=1}^{n} Y_{ijkl}^2 - \text{CF}$$

$$= (54)^2 + (52)^2 + (41)^2 + \ldots + (77)^2 - \text{CF}$$

$$= 67304.00 - 64516.00 = 2788.00$$

The sum of squares of the main effects is computed from Eqs. (4.25) to (4.27).

$$SS_A = \frac{\sum_{i=1}^{a} T_{i...}^2}{bcn} - \text{CF}$$

$$= \frac{(423)^2 + (593)^2}{8} - \text{CF}$$

$$= 66322.25 - 64516.00 = 1806.25$$

$$SS_B = \frac{\sum_{j=1}^{b} T_{.j..}^2}{acn} - \text{CF}$$

$$= \frac{(499)^2 + (517)^2}{8} - \text{CF}$$

$$= 64536.25 - 64516.00 = 20.25$$

The level totals for factor $C(T_{..k.})$ are:

Level 1 totals = 274 + 277 = 551
Level 2 totals = 225 + 240 = 465

$$SS_C = \frac{\sum_{k=1}^{c} T_{..k.}^2}{abn} - \text{CF}$$

$$= \frac{(551)^2 + (465)^2}{8} - \text{CF}$$

$$= 64978.25 - 64516.00 = 462.25$$

To compute the two-factor interaction sum of squares, we have to identify the respective two factors combined totals.

Totals for *AB* interaction sum of squares are found from

$$A_1B_1 = 106 + 99 = 205$$
$$A_1B_2 = 168 + 126 = 294$$
$$A_2B_1 = 120 + 98 = 218$$
$$A_2B_2 = 157 + 142 = 299$$

Note that in each *AB* total there are 4 observations.
Sum of squares for *AB* interaction is found from Eq. (4.28).

$$SS_{AB} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b} T_{ij..}^2}{cn} - \text{CF} - SS_A - SS_B$$

$$= \frac{(205)^2 + (294)^2 + (218)^2 + (299)^2}{4} - \text{CF} - SS_A - SS_B$$

$$= 66346.50 - 64516.00 - 1806.25 - 20.25 = 4.00$$

Totals for *AC* interaction sum of squares are found from

$$A_1C_1 = 106 + 120 = 226$$
$$A_1C_2 = 99 + 98 = 197$$
$$A_2C_1 = 168 + 157 = 325$$
$$A_2C_2 = 126+ 142 = 268$$

Note that in each *AC* total also there are 4 observations.
Sum of squares for *AC* interaction is found from Eq. (4.29).

$$SS_{AC} = \frac{\sum_{i=1}^{a}\sum_{k=1}^{c} T_{i.k.}^2}{bn} - \text{CF} - SS_A - SS_C$$

$$= \frac{(226)^2 + (197)^2 + (325)^2 + (268)^2}{4} - \text{CF} - SS_A - SS_C$$

$$= 66833.50 - 64516.00 - 1806.25 - 462.25 = 49.00$$

The *BC* interaction totals ($Y_{.jk.}$) are given in Table 4.6. The sum of squares for *AB* is computed from Eq. (4.30).

$$SS_{BC} = \frac{\sum_{j=1}^{b}\sum_{k=1}^{c} T_{.jk.}^2}{an} - \text{CF} - SS_B - SS_C$$

$$= \frac{(274)^2 + (225)^2 + (277)^2 + (240)^2}{4} - \text{CF} - SS_B - SS_C$$

$$= 65007.50 - 64516.00 - 20.25 - 462.25 = 9.00$$

The three-factor interaction $ABC$ is found using the $ABC$ cell totals (underlined numbers in Table 4.6). It is computed from Eq. (4.31).

$$SS_{ABC} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} T_{ijk.}^2}{n} - \text{CF} - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC}$$

$$= \frac{(106)^2 + (99)^2 + (120)^2 + \ldots + (142)^2}{2} - \text{CF} - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC}$$

$$= 66977.00 - 64516.00 - 1806.25 - 20.25 - 462.25 - 4.00 - 49.00 - 9.00$$

$$= 110.25$$

$$SS_e = SS_{\text{Total}} - \text{SS of all effects}$$

$$= 2788.00 - 2461.00 = 327.00$$

All these computations are summarized in the ANOVA (Table 4.7).

**TABLE 4.7** Analysis of variance for Illustration 4.2

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean squares* | $F_0$ |
|---|---|---|---|---|
| Depth of cut ($A$) | 1806.25 | 1 | 1806.25 | 44.19 |
| Speed ($B$) | 20.25 | 1 | 20.25 | 0.495 |
| Feed ($C$) | 462.25 | 1 | 462.25 | 11.31 |
| $AB$ | 4.00 | 1 | 4.00 | 0.098 |
| $AC$ | 49.00 | 1 | 49.00 | 1.199 |
| $BC$ | 9.00 | 1 | 9.00 | 0.22 |
| $ABC$ | 110.25 | 1 | 110.25 | 2.697 |
| Pure error | 327.00 | 8 | 40.875 | — |
| Total | 2788.00 | 15 | | |

At 5% significance level, $F_{0.05,1,8} = 5.32$. So, the main effects depth of cut and feed alone are significant. That is, the feed and depth of cut influence the surface finish.

The procedure followed in the analysis of three-factor factorial design can be extended to any number of factors and also factors with more than two levels. However with the increase in problem size manual computation would be difficult and hence a computer software can be used.

## 4.4 RANDOMIZED BLOCK FACTORIAL EXPERIMENTS

We have already discussed the application of blocking principle in a single-factor experiment. Blocking can be incorporated in a factorial design also. Consider a two-factor factorial design with factors $A$ and $B$ and $n$ replicates. The statistical model for this design is [Eq. (4.1)] is as follows:

$$Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + e_{(ij)k} \qquad \begin{cases} i = 1, 2, ..., a \\ j = 1, 2, ..., b \\ k = 1, 2, ..., n \end{cases}$$

The analysis of this design is already discussed earlier. Suppose to run this experiment a particular raw material is required that is available in small batches. And each batch of material is just enough to run one complete replication only (*abn* treatments). Then for replication another batch is needed. This causes the experimental units to be non-homogeneous. An alternative design is to run each replicate with each batch of material which is a randomization restriction or blocking. This type of design is called *randomized block factorial design*. The order of experimentation is random with in the block. Similarly if an experiment with two replications could not be completed in one day, we can run one replication on one day and the second replication on another day. Thus, in this, day represents the block.

### 4.4.1 The Statistical Model

$$Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \delta_k + e_{(ij)k} \qquad \begin{cases} i = 1, 2, ..., a \\ j = 1, 2, ..., b \\ k = 1, 2, ..., n \end{cases} \tag{4.33}$$

where, $\delta_k$ is the effect of *k*th block.

The model assumes that interaction between blocks and treatments/factors is negligible. The analysis of this model is same as that of factorial design except that the error sum of squares is partitioned into block sum of squares and error sum of squares. Block sum of squares is computed using the block totals ($T_{..k}$). A typical ANOVA table for a two-factor block factorial design is given in Table 4.8.

**TABLE 4.8** General ANOVA for a two-factor block factorial design

| *Source of variation* | *Sum of squares* | *Degrees of freedom* |
|---|---|---|
| Blocks | $\frac{1}{ab} \sum_k T_{..k}^2 - \text{CF}$ | $(n - 1)$ |
| $A$ | $\frac{1}{bn} \sum_i T_{i..}^2 - \text{CF}$ | $(a - 1)$ |
| $B$ | $\frac{1}{an} \sum_j T_{.j.}^2 - \text{CF}$ | $(b - 1)$ |
| $AB$ | $\frac{1}{n} \sum_i \sum_j T_{ij.}^2 - \text{CF} - SS_A - SS_B$ | $(a - 1)(b - 1)$ |
| Error | By subtraction | |
| Total | $\sum_i \sum_j \sum_k Y_{ijk}^2 - \text{CF}$ | $abn - 1$ |

The Correction factor in Table 4.8 is $\dfrac{T_{...}^2}{abn}$ ($T_{...}$ = Grand total).

## ILLUSTRATION 4.3

**Randomized Block Factorial Design**

A study was conducted to investigate the effect of frequency of lifting (number of lifts/min) and the amount of load (kg) lifted from floor level to a knee level height (51 cm). The procedure consists of lifting a standard container with the given load at a given frequency. Three loads 5, 10 and 15 kg and three frequencies 2, 4 and 6 lifts per minute have been studied. The percentage rise in heart rate was used as response. Four workers have been selected and used in the study. A randomized block factorial design was used with load and frequency as factors and workers as blocks. The data are given in Table 4.9.

**TABLE 4.9** Percentage rise in heart rate data for Illustration 4.3

| *Workers-Block* ($\delta_i$) | Frequency ($A_j$) 2 | | | 4 | | | 6 | | | $T_{i..}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | Load ($B_k$) | | | Load ($B_k$) | | | Load ($B_k$) | | | |
| | 5 | 10 | 15 | 5 | 10 | 15 | 5 | 10 | 15 | |
| 1 | 5.2 | 8.6 | 13.5 | 5.2 | 10.0 | 14.5 | 5.5 | 12.4 | 16.0 | 90.9 |
| 2 | 8.5 | 19.2 | 22.6 | 8.9 | 16.2 | 27.5 | 13.0 | 29.3 | 48.0 | 193.2 |
| 3 | 5.4 | 4.4 | 5.8 | 7.0 | 7.5 | 13.7 | 10.2 | 17.1 | 24.5 | 95.6 |
| 4 | 3.2 | 4.3 | 5.9 | 5.4 | 9.4 | 19.7 | 7.9 | 15.2 | 21.3 | 92.3 |
| $T_{jk.}$ | 22.3 | 36.5 | 47.8 | 26.5 | 43.1 | 75.4 | 36.6 | 74.0 | 109.8 | $T_{...}$ = 472.0 |
| $T_{j..}$ | | 106.6 | | | 145.0 | | | 220.4 | | |

***Data analysis:*** Let factor $A$ = frequency, factor $B$ = load and the block = $\delta$.

The data is analyzed using ANOVA. All the relevant totals required for computing sum of squares is given in Table 4.9.

$$\text{Correction factor (CF)} = \frac{T_{...}^2}{abn} = \frac{(472)^2}{36} = 6188.444$$

$$SS_{\text{Total}} = \sum_{i=1}^{4} \sum_{j=1}^{3} \sum_{k=1}^{3} Y_{ijk}^2 - \text{CF}$$

$$= (5.2)^2 + (8.6)^2 + (13.5)^2 + \cdots + (21.3)^2 - 6188.444$$

$$= 2941.236$$

The sum of squares of main effects is computed using the respective factor level total.

The sum of squares of $A$ is computed from ($T_{j..}$)

$$SS_A = \frac{(106.6)^2 + (145.0)^2 + (220.4)^2}{12} - \text{CF} = 558.616$$

For factor $B$, the level totals are:

$$B_5 = 22.3 + 26.5 + 36.6 = 85.4$$

$$B_{10} = 36.5 + 43.1 + 74.0 = 153.6$$

$$B_{15} = 47.8 + 75.4 + 109.8 = 233.0$$

$$SS_B = \frac{(85.4)^2 + (153.6)^2 + (233.0)^2}{12} - \text{CF} = 909.483$$

The interaction between frequency and load ($AB$) is found using the combined totals ($T_{.jk.}$).

$$SS_{AB} = \frac{(22.3)^2 + (36.5)^2 + \cdots + (109.8)^2}{4} - \text{CF} - SS_A - SS_B$$

$$= 7807.5 - 6188.444 - 558.616 - 909.483 = 151.207$$

The block sum of squares is found from the row totals ($T_{i..}$ ).

$$SS_{\text{Block}} = \frac{(90.9)^2 + (193.2)^2 + (95.6)^2 + (92.3)^2}{9} - \text{CF}$$

$$= 7027.522 - 6188.444 = 839.078$$

Table 4.10 gives the ANOVA for Illustration 4.3.

**TABLE 4.10** ANOVA for Illustration 4.3 (RBFD)

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean squares* | $F_0$ |
|---|---|---|---|---|
| Workers (Blocks) | 839.08 | 3 | — | — |
| Frequency ($A$) | 558.62 | 2 | 279.31 | 13.88 |
| Load ($B$) | 909.48 | 2 | 454.74 | 22.60 |
| Frequency × Load ($AB$) | 151.21 | 4 | 37.80 | 1.88 |
| Error | 482.85 | 24 | 20.12 | |
| Total | 2941.24 | 35 | | |

$F_{0.05,2,24} = 3.40$, $F_{0.05,4,24} = 2.78$

***Inference:*** At 5% level of significance, frequency and load have got significant effect on the percentage rise in heart rate due to the lifting task. However, frequency × load interaction shows no significance.

## 4.5 EXPERIMENTS WITH RANDOM FACTORS

In all the experiments discussed so far we have considered only fixed factors (the levels of the factors are fixed or chosen specifically). The model that describes a fixed factor experiment is called a *fixed effects model*. Here, we consider the treatment effects $T_i$s as fixed constants. The null hypothesis to be tested is

$$H_0\text{: } T_i = 0 \qquad \text{for all } i \tag{4.34}$$

And our interest here is to estimate the treatment effects. In these experiments, the conclusions drawn are applicable only to the levels considered in the experimentation. For testing the fixed effect model, note that we have used mean square error ($MS_e$) to determine the $F$-statistic of all the effects in the model (main effects as well as interaction effects).

### 4.5.1 Random Effects Model

If the levels of the factors are chosen randomly from a population, the model that describes such factors is called a *random effects model*. In this case the $T_i$s are considered as *random variable*s and are NID (0, $\sigma_T^2$). $\sigma_T^2$ represents the variance among the $T_i$s. The null hypothesis here is

$$H_0\text{: } \sigma_T^2 = 0 \tag{4.35}$$

In these experiments we wish to estimate $\sigma_T^2$ (variance of $T_i$s). Since the levels are chosen randomly, the conclusions drawn from experimentation can be extended to the entire population from which the levels are selected.

The main difference between fixed factor experiment and the random factor experiment is that in the later case the denominator (mean square) used to compute the $F$-statistic in the ANOVA to test any effect has to be determined by deriving the Expected Mean Square (EMS) of the effects. Similarly, EMS has to be derived in the case of mixed effects model (some factors are fixed and some are random) to determine the appropriate $F$-statistic.

Suppose we have a factorial model with two factors $A$ and $B$. The effects under the three categories of models would be as follows:

| *Fixed effects model* | *Random effects model* | *Mixed effects model* |
|---|---|---|
| $A$ is fixed | $A$ is random | $A$ is fixed |
| $B$ is fixed | $B$ is random | $B$ is random |
| $AB$ is also fixed | $AB$ is random | $AB$ will be random |

### 4.5.2 Determining Expected Mean Squares

The experimental data is analyzed using ANOVA. This involves the determination of sum of squares and degrees of freedom of all the effects and finding appropriate test statistic for each effect. This requires the determination of EMS while we deal with random or mixed effects models. For this purpose the following procedure is followed:

1. The error term in the model $e_{ij\ldots m}$ shall be written as $e_{(ij\ldots)m}$, where the subscript $m$ denotes the replication. That is, the subscript representing the replication is written outside the parenthesis. This indicates that $(ij_{\ldots})$ combination replicated $m$ times.

2. Subscripts within the parentheses are called *dead subscripts* and the subscript outside the parenthesis is the *live subscript*.
3. The degrees of freedom for any term in the model is the product of the number of levels associated with each dead subscript and the number of levels minus one associated with each live subscript.

   For example, the degrees of freedom for $(AB)_{ij}$ is $(a - 1)(b - 1)$ and the degrees of freedom for the term $e_{(ij)k}$ is $ab(n - 1)$.
4. The fixed effect (fixed factor) of any factor $A$ is represented by $\phi_A$, where $\phi_A = \dfrac{\sum T_i^2}{a-1}$ and the random effect (random factor) of $A$ is represented by $\sigma_A^2$. If an interaction contains at least one random effect, the entire interaction is considered as random.

#### *Deriving expected mean square*

Consider a two-factor fixed effect model with factors $A$ and $B$.

$$Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + e_{(ij)k} \quad \begin{cases} i = 1, 2, ..., a \\ j = 1, 2, ..., b \\ k = 1, 2, ..., n \end{cases}$$

In this model $\mu$ is a constant term and all other terms are variable.

*Step* 1: Write down the variable terms as row headings.

| |
|---|
| $A_i$ |
| $B_j$ |
| $(AB)_{ij}$ |
| $e_{(ij)k}$ |

*Step* 2: Write the subscripts as column headings.

| | $i$ | $j$ | $k$ | |
|---|---|---|---|---|
| $A_i$ | | | | |
| $B_j$ | | | | |
| $(AB)_{ij}$ | | | | |
| $e_{(ij)k}$ | | | | |

*Step* 3: Above the column subscripts, write $F$ if the subscript belongs to a fixed factor and $R$ if the subscript belongs to a random factor. Also, write the number of levels/observations above each subscript.

| | $a$<br>$F$<br>$i$ | $b$<br>$F$<br>$j$ | $n$<br>$R$<br>$k$ | |
|---|---|---|---|---|
| $A_i$ | | | | |
| $B_j$ | | | | |
| $(AB)_{ij}$ | | | | |
| $e_{(ij)k}$ | | | | |

Note that the replication is always random.

*Step* 4: Fill the subscript matching cells with 0 if the column subscript is fixed and 1 if the column subscript is random and also by 1 if the subscripts are dead.

| | $a$<br>$F$<br>$i$ | $b$<br>$F$<br>$j$ | $n$<br>$R$<br>$k$ | |
|---|---|---|---|---|
| $A_i$ | 0 | | | |
| $B_j$ | | 0 | | |
| $(AB)_{ij}$ | 0 | 0 | | |
| $e_{(ij)k}$ | 1 | 1 | 1 | |

*Step* 5: Fill in the remaining cells with the number of levels shown above the column headings.

| | $a$<br>$F$<br>$i$ | $b$<br>$F$<br>$j$ | $n$<br>$R$<br>$k$ | EMS |
|---|---|---|---|---|
| $A_i$ | 0 | $b$ | $n$ | |
| $B_j$ | $a$ | 0 | $n$ | |
| $(AB)_{ij}$ | 0 | 0 | $n$ | |
| $e_{(ij)k}$ | 1 | 1 | 1 | |

*Step 6*: Obtain EMS explained as follows:

Consider the row effect $A_i$ and suppress the column(s) corresponding to this subscript ($i$th column).

Take the product of the visible row numbers ($bn$) and the corresponding effect ($\phi_A$) and enter in that row under the EMS column which is $bn\ \phi_A$. This is repeated for all the effects, wherever this subscript ($i$) appears. The next rows where this subscript appears are in $(AB)_{ij}$ and $e_{(ij)k}$. The EMS terms corresponding to these two terms are zero ($0 \times n$) and $\sigma_e^2$(variance associated with the error term) respectively. These terms are added to the first term, $bn\ \phi_A$. Thus, the EMS of $A_i$ is $bn\ \phi_A + \sigma_e^2$ (zero is omitted). Note that the EMS for the error term is always $\sigma_e^2$. This procedure is repeated for all the effects. Table 4.11 gives the EMS for the two-factor fixed effects model and $F$-statistic.

**TABLE 4.11** Expected mean squares for a two-factor fixed effects model

| | $a$<br>$F$<br>$i$ | $b$<br>$F$<br>$j$ | $n$<br>$R$<br>$k$ | EMS | $F_0$ |
|---|---|---|---|---|---|
| $A_i$ | 0 | $b$ | $n$ | $nb\ \phi_A + \ \sigma_e^2$ | $MS_A/MS_e$ |
| $B_j$ | $a$ | 0 | $n$ | $na\ \phi_B + \ \sigma_e^2$ | $MS_B/MS_e$ |
| $(AB)_{ij}$ | 0 | 0 | $n$ | $n\ \phi_{AB} + \ \sigma_e^2$ | $MS_{AB}/MS_e$ |
| $e_{(ij)\,k}$ | 1 | 1 | 1 | $\sigma_e^2$ | |

The $F$-statistic is obtained by forming a ratio of two expected mean squares such that the only term in the numerator that is not in the denominator is the effect to be tested. Thus, the $F$-statistic (Table 4.11) for all the effects is obtained by dividing the mean square of the effect with the error mean square. This will be the case in all the fixed effect models.

Tables 4.12, 4.13 and 4.14 give the expected mean squares for a two-factor random effects model, a two-factor mixed effects model and a three-factor random effects model respectively.

**TABLE 4.12** Expected mean squares for a two-factor random effects model

| | $a$<br>$R$<br>$i$ | $b$<br>$R$<br>$j$ | $n$<br>$R$<br>$k$ | EMS | $F_0$ |
|---|---|---|---|---|---|
| $A_i$ | 1 | $b$ | $n$ | $nb\ \sigma_A^2 + n\ \sigma_{AB}^2\ \sigma_e^2$ | $MS_A/MS_{AB}$ |
| $B_j$ | $a$ | 1 | $n$ | $na\ \sigma_B^2 + n\ \sigma_{AB}^2\ \sigma_e^2$ | $MS_B/MS_{AB}$ |
| $(AB)_{ij}$ | 1 | 1 | $n$ | $n\ \sigma_{AB}^2 + \ \sigma_e^2$ | $MS_{AB}/MS_e$ |
| $e_{(ij)\,k}$ | 1 | 1 | 1 | $\sigma_e^2$ | |

**TABLE 4.13** Expected mean squares for a two-factor mixed effects model

| | $a$<br>$F$<br>$i$ | $b$<br>$R$<br>$j$ | $n$<br>$R$<br>$k$ | EMS | $F_0$ |
|---|---|---|---|---|---|
| $A_i$ | 0 | $b$ | $n$ | $nb\ \phi_A + n\sigma_{AB}^2 + \sigma_e^2$ | $MS_A/MS_{AB}$ |
| $B_j$ | $a$ | 1 | $n$ | $na\ \sigma_B^2 + \ \sigma_e^2$ | $MS_B/MS_e$ |
| $(AB)_{ij}$ | 0 | 1 | $n$ | $n\ \sigma_{AB}^2 + \ \sigma_e^2$ | $MS_{AB}/MS_e$ |
| $e_{(ij)\,k}$ | 1 | 1 | 1 | $\sigma_e^2$ | |

**TABLE 4.14** Expected mean squares for a three-factor random effects model

| | $a$<br>$R$<br>$i$ | $b$<br>$R$<br>$j$ | $c$<br>$R$<br>$k$ | $n$<br>$R$<br>$l$ | EMS |
|---|---|---|---|---|---|
| $A_i$ | 1 | $b$ | $c$ | $n$ | $bcn\,\sigma_A^2 + cn\,\sigma_{AB}^2 + bn\,\sigma_{AC}^2 + n\sigma_{ABC}^2 + \sigma_e^2$ |
| $B_j$ | $a$ | 1 | $c$ | $n$ | $acn\,\sigma_B^2 + cn\,\sigma_{AB}^2 + an\,\sigma_{BC}^2 + n\sigma_{ABC}^2 + \sigma_e^2$ |
| $C_k$ | $a$ | $b$ | 1 | $n$ | $abn\,\sigma_C^2 + an\,\sigma_{BC}^2 + bn\,\sigma_{AC}^2 + n\,\sigma_{ABC}^2 + \sigma_e^2$ |
| $(AB)_{ij}$ | 1 | 1 | $c$ | $n$ | $cn\,\sigma_{AB}^2 + n\,\sigma_{ABC}^2 + \sigma_e^2$ |
| $(AC)_{ik}$ | $a$ | 1 | 1 | $n$ | $bn\,\sigma_{AC}^2 + n\,\sigma_{ABC}^2 + \sigma_e^2$ |
| $(BC)_{jk}$ | 1 | $b$ | 1 | $n$ | $an\,\sigma_{BC}^2 + n\,\sigma_{ABC}^2 + \sigma_e^2$ |
| $(ABC)_{ijk}$ | 1 | 1 | 1 | $n$ | $n\,\sigma_{ABC}^2 + \sigma_e^2$ |
| $e_{(ijk)l}$ | 1 | 1 | 1 | 1 | $\sigma_e^2$ |

Note that when all the three factors are random, the main effects cannot be tested.

### 4.5.3 The Approximate *F*-test

In factorial designs with three or more factors involving random or mixed effects model and certain complex designs, we may not be able to find exact test statistic for certain effects in the model. These may include main effects also. Under these circumstances we may adopt one of the following approaches.

1. Assume certain interactions are negligible. This may facilitate the testing of those effects for which exact statistic is not available. For example, in a three-factor random effects model (Table 4.14), if we assume interaction $AC$ as zero, the main effects $A$ and $C$ can be tested. Unless some prior knowledge on these interactions is available, the assumption may lead to wrong conclusions.
2. The second approach is to test the interactions first and set the insignificant interaction (s) to zero. Then the other effects are tested assuming the insignificant interactions as zero.
3. Another approach is to pool the sum of squares of some effects to obtain more degrees of freedom to the error term that enable to test other effects. It is suggested to pool the sum of squares of those effects for which the $F$-statistic is not significant at a large value of $\alpha$ (0.25). It is also recommended to pool only when the error sum of squares has less than six degrees of freedom.

## 4.6 RULES FOR DERIVING DEGREES OF FREEDOM AND SUM OF SQUARES

To determine the degrees of freedom and sum of squares for any term in the statical model of the experiment, we use the following rules.

### 4.6.1 Rule for Degrees of Freedom

Number of degrees of freedom for any term in the model = (Number of levels of each dead subscript) × (Number of levels – 1) of each live subscript (4.36)

Suppose we have the following model:

$$Y_{ijkl} = \mu + A_i + B_j + (AB)_{ij} + C_{k(j)} + (AC)_{ik(j)} + e_{(ijk)l} \quad \begin{cases} i = 1, 2, ..., a \\ j = 1, 2, ..., b \\ k = 1, 2, ..., c \\ l = 1, 2, ..., n \end{cases} \tag{4.37}$$

Degrees of freedom for $AC$ interaction = $b\ (a - 1)\ (c - 1)$ (4.38)

Similarly, the error degrees of freedom = $abc\ (n - 1)$ (4.39)

### 4.6.2 Rule for Computing Sum of Squares

Suppose we want to compute sum of squares for the term $AC_{ik(j)}$.

Define 1 = CF (Correction factor)

Write the degrees of freedom (df) for $AC$ and expand algebraically

df for $AC = b(a - 1)\ (c - 1)$

On expansion, we obtain $abc - bc - ab + b$ (4.40)

In Eq. (4.40), if 1 is present, it will be the correction factor.

Now we write the sum of squares as follows using (.) notation.

$$SS_{AC} = \sum_i \sum_j \sum_k \frac{T_{ijk.}^2}{n} - \sum_j \sum_k \frac{T_{.jk.}^2}{an} - \sum_i \sum_j \frac{T_{ij..}^2}{cn} + \sum_j \frac{T_{.j..}^2}{acn} \tag{4.41}$$

The first term in Eq. (4.40) is $abc$. So to write SS for this term, triple summation over $ijk$ corresponding to the levels of $abc$ is written and its total ($T_{ijk.}^2$) is written keeping dot (.) in the place of subscript $l$. This subscript belongs to the replication ($n$) and it becomes the denominator for the first term of Eq. (4.41). Similarly, all other terms are obtained.

## PROBLEMS

**4.1** A study was conducted to investigate the effect of feed and cutting speed on the power consumption (W) of a drilling operation. The data is given in Table 4.15.

**TABLE 4.15** Data for Problem 4.1

| *Cutting speed* (m/min) | *Feed* (mm/rev) 0.05 | 0.10 | 0.20 |
|---|---|---|---|
| 30 | 17.40 | 26.35 | 38.00 |
| | 17.50 | 26.30 | 38.20 |
| 40 | 23.60 | 35.10 | 54.60 |
| | 23.40 | 35.15 | 54.85 |
| 50 | 29.30 | 39.05 | 58.50 |
| | 29.25 | 38.90 | 58.30 |

Analyse the data and draw the conclusions. Use $\alpha = 0.05$.

**4.2** An experiment was conducted to study the effect of type of tool and depth of cut on the power consumption. Data obtained are given in Table 4.16.

**TABLE 4.16** Data for Problem 4.2

| *Type of tool* (m/min) | *Depth of cut* (mm/rev) 1.0 | 2.0 | 3.0 |
|---|---|---|---|
| 1 | 5.0 | 11.1 | 17.0 |
| | 4.8 | 11.3 | 17.3 |
| 2 | 5.2 | 12.1 | 18.3 |
| | 5.4 | 12.3 | 18.1 |

Analyse the data and draw the conclusions. Use $\alpha = 0.05$.

**4.3** The bonding strength of an adhesive has been studied as a function of temperature and pressure. Data obtained are given in Table 4.17.

**TABLE 4.17** Data for Problem 4.3

| *Pressure* | *Temperature* 30 | 60 | 90 |
|---|---|---|---|
| 120 | 90 | 34 | 58 |
| | 74 | 80 | 30 |
| 130 | 150 | 106 | 35 |
| | 159 | 115 | 70 |
| 140 | 138 | 110 | 96 |
| | 168 | 160 | 104 |

(a) Analyse the data and draw conclusions. Use $\alpha = 0.05$.
(b) Prepare appropriate residual plots and comment on the model adequacy.
(c) Under what conditions the process should be operated. Why?

**4.4** A study was conducted to investigate the effect of frequency and load lifted from knee to waist level. Three frequencies (number of lifts per minute) and three loads of lift (kg) have been studied. Four subjects (workers) involved in manual lifting operation were studied. The percentage of rise in heart rate measured for each subject (worker) is given in Table 4.18.

**TABLE 4.18** Data for Problem 4.4

| *Subject (Block)* | *Frequency* 2 | | | 4 | | | 6 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Load | | | Load | | | Load | | |
| | 5 | 10 | 15 | 5 | 10 | 15 | 5 | 10 | 15 |
| 1 | 2.61 | 4.47 | 5.27 | 4.26 | 10.56 | 16.77 | 4.52 | 10.02 | 15.24 |
| | 4.49 | 4.80 | 4.86 | 7.88 | 12.15 | 15.23 | 9.06 | 13.10 | 20.92 |
| 2 | 3.74 | 6.02 | 19.63 | 5.28 | 9.95 | 21.98 | 9.68 | 20.76 | 43.51 |
| | 3.52 | 5.95 | 12.43 | 4.41 | 13.91 | 16.15 | 3.88 | 27.26 | 40.38 |
| 3 | 2.29 | 2.87 | 7.92 | 3.96 | 11.79 | 17.91 | 10.51 | 14.11 | 24.15 |
| | 1.64 | 3.54 | 10.49 | 4.78 | 7.92 | 15.98 | 4.99 | 11.63 | 30.18 |
| 4 | 4.30 | 9.26 | 11.03 | 9.64 | 25.89 | 27.23 | 6.59 | 25.49 | 39.99 |
| | 3.44 | 14.55 | 13.53 | 3.63 | 22.01 | 26.86 | 15.23 | 24.86 | 37.73 |

Analyse the data and draw the conclusions. Use $\alpha = 0.05$. Note that it is a randomized block factorial design.

**4.5** A study was conducted to investigate the effect of type of container on the heart rate of manual materials lifting workers. Two types of containers (rectangular box and bucket) have been used in the study to lift 10 kg, 15 kg and 20 kg of load from floor to 51 cm height level. Five workers were selected randomly and collected data (heart rate) which is given in Table 4.19. The order of experimentation was random with in each block.

**TABLE 4.19** Data for Problem 4.5

| *Subject (Block)* | *Type of container* *Rectangular box* | | | *Bucket* | | |
|---|---|---|---|---|---|---|
| | Load | | | Load | | |
| | 10 | 15 | 20 | 10 | 15 | 20 |
| 1 | 76 | 72 | 88 | 68 | 80 | 68 |
| | 76 | 72 | 88 | 68 | 72 | 68 |
| 2 | 76 | 84 | 100 | 88 | 88 | 76 |
| | 76 | 84 | 100 | 92 | 88 | 80 |
| 3 | 92 | 100 | 104 | 80 | 92 | 80 |
| | 92 | 104 | 100 | 76 | 88 | 84 |
| 4 | 84 | 88 | 92 | 80 | 72 | 80 |
| | 88 | 90 | 94 | 80 | 72 | 82 |

Analyse the data and draw the conclusions. Use $\alpha = 0.05$.

**4.6** Three different methods of gray component replacement in a printing process have been studied. A test colour patch in two different types of paper was printed using two different types of digital printing machines. The colour difference obtained from each method was measured (Table 4.20). Analyse the data and draw conclusions. Use $\alpha = 0.05$.

**TABLE 4.20** Data for Problem 4.6

| *Method* | *Type of printing machine* | | | |
|---|---|---|---|---|
| | *HP* | | *Xerox* | |
| | *Paper* | | *Paper* | |
| | *Coated* | *Uncoated* | *Coated* | *Uncoated* |
| PCS | 3.08 | 7.35 | 9.52 | 16.81 |
| | 3.10 | 7.30 | 9.25 | 16.50 |
| | 3.05 | 7.50 | 9.10 | 17.01 |
| Device link | 2.85 | 4.51 | 7.43 | 15.26 |
| | 2.87 | 4.20 | 7.41 | 15.50 |
| | 2.83 | 4.61 | 7.40 | 15.05 |
| Dynamic device link | 2.46 | 3.71 | 7.41 | 9.78 |
| | 2.40 | 3.80 | 7.55 | 9.85 |
| | 2.39 | 3.65 | 7.77 | 9.99 |

**4.7** An industrial engineer has conducted a study to investigate the effect of type of fixture, operator and layout of components on the assembly time of a product. The operators were selected randomly. The assembling time data obtained is given in Table 4.21. Analyse the data and draw conclusions.

**TABLE 4.21** Data for Problem 4.7

| *Fixture* | *Type of layout* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | | | 2 | | | 3 | | |
| | *Operator* | | | *Operator* | | | *Operator* | | |
| | $O_1$ | $O_2$ | $O_3$ | $O_1$ | $O_2$ | $O_3$ | $O_1$ | $O_2$ | $O_3$ |
| 1 | 5.2 | 10.1 | 14.5 | 4.3 | 10.5 | 16.7 | 3.2 | 9.6 | 12.3 |
| | 5.7 | 11.9 | 11.6 | 7.9 | 12.2 | 15.2 | 3.5 | 7.7 | 12.9 |
| 2 | 8.9 | 16.2 | 27.4 | 5.2 | 9.9 | 21.0 | 12.6 | 24.3 | 30.5 |
| | 14.3 | 19.7 | 23.0 | 4.4 | 13.9 | 16.2 | 18.8 | 25.2 | 30.5 |
| 3 | 7.1 | 7.5 | 13.7 | 4.0 | 11.2 | 17.9 | 10.7 | 11.9 | 17.6 |
| | 5.4 | 8.4 | 16.9 | 4.7 | 7.9 | 27.0 | 9.9 | 11.2 | 16.2 |
| 4 | 5.4 | 9.4 | 19.7 | 9.6 | 25.2 | 27.2 | 13.5 | 26.8 | 32.5 |
| | 4.6 | 9.6 | 17.4 | 6.0 | 22.0 | 26.8 | 13.9 | 27.7 | 36.4 |

CHAPTER 5

# The $2^k$ Factorial Experiments

## 5.1 INTRODUCTION

Factorial designs with several factors are common in research, especially at the early stages of process/product development. When several factors are to be studied, the number of levels of these factors is to be limited; otherwise the number of experiments would be very large. And each factor with a minimum of two levels has to be studied in order to obtain the effect of a factor. Thus, several factors with two levels each will minimize the number of experiments.

A factorial design with $k$-factors, each at only two levels is called $2^k$ design, a special case of multi-factor factorial experiment. It provides the smallest number of treatment combinations with which $k$-factors can be studied in a full factorial experiment. These designs are usually used as factor screening experiments. When several factors are involved, all of them may not affect the process or product performance. The $2^k$ design assists the experimenter to identify the few significant factors among them. With these few significant factors, full factorial experiment with more levels can then be studied to determine the optimal levels for the factors.

Since there are only two levels for each factor, we assume that the response is approximately linear over the levels of the factors. The levels of the factors may be quantitative or qualitative.

*Quantitative:* two values of temperature, pressure or time etc.
*Qualitative:* two machines, two operators, the high or low levels of a factor, etc.

Generally, the following assumptions are made in the $2^k$ factorial designs.

- The factors are fixed.
- The designs are completely randomized.
- The usual normality assumption holds good.

## 5.2 THE $2^2$ FACTORIAL DESIGN

The simplest design in the $2^k$ series is a $2^2$ design; two factors each at two levels. The two levels of the factors are usually referred to as *low* and *high*. Let us study this design with an example.

**ILLUSTRATION 5.1**

**The $2^2$ Factorial Experiment**

Consider the problem of studying the effect of temperature and pressure on the yield of a chemical process. Let the two factors temperature and pressure are denoted by $A$ and $B$ respectively. Suppose the following results are obtained from this study. The number of replications is two.

| *Pressure* ($B$) | *Temperature* ($A$) | |
|---|---|---|
| | *Low* (60°C) | *High* (70°C) |
| Low | 40 | 43 |
| (100 MPa) | 37 | 50 |
| High | 59 | 37 |
| (150 MPa) | 54 | 43 |

This data can be analysed using the method similar to that of the two-factor factorial experiment discussed in Chapter 4. This method becomes cumbersome when the number of factors increases. Hence, a different method is adopted for $2^k$ design, which is discussed below.

The response (yield) total from the two replications for the four-treatment combinations is shown below.

| *Treatment combination* | *Replicate* | | |
|---|---|---|---|
| | 1 | 2 | *Total* |
| $A$ low, $B$ low | 40 | 37 | 77 |
| $A$ high, $B$ low | 43 | 50 | 93 |
| $A$ low, $B$ high | 59 | 54 | 113 |
| $A$ high, $B$ high | 37 | 43 | 80 |
| | | Grand total = | 363 |

The treatment combinations with the yield data are shown graphically in Figure 5.1. By convention, the effect of a factor is denoted by a capital letter and the treatment combinations are represented by lower case letters.

Thus, $A$ refers to the effect of $A$
$B$ refers to the effect of $B$
$AB$ refers to the effect of $AB$ interaction.

For any factor, the low level is represented by 0 or – or 1. And the high level of the factor is represented by 1 or + or 2. Thus, we have three notations, namely (0, 1), (–, +) and (1, 2) in these $2^k$ designs.

Under (0, 1) notation, the four treatment combinations are represented as follows:

0 0: Both factors at low level; denoted by (1)
1 0: $A$ high, $B$ low; denoted by $a$
0 1: $A$ low, $B$ high; denoted by $b$
11: $A$ high, $B$ high: denoted by $ab$

These four treatment combinations can be represented by the coordinates of the vertices of the square as shown in Figure 5.1.

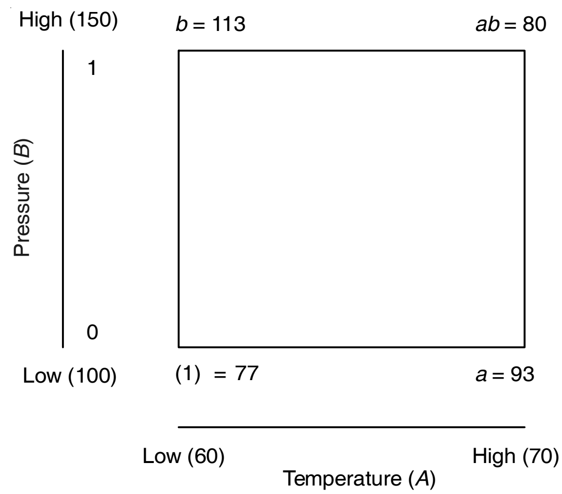



**FIGURE 5.1** Graphical representation of $2^2$ design.

### 5.2.1 Determining the Factor Effects

In a two-level design, the average effect of a factor is defined as the change in response produced due to change in the level of that factor averaged over the other factor. In Figure 5.1, (1), $a$, $b$ and $ab$ represent the response total of all the $n$ replicates of the treatment combinations.

***Main effect of A:***

$$\text{Effect of } A \text{ at the low level of factor } B = \frac{[a - (1)]}{n}$$

$$\text{Effect of } A \text{ at the high level of factor } B = \frac{(ab - b)}{n}$$

$$\begin{aligned} \text{Average effect of } A &= 1/2n\ [(a - 1) + (ab - b)] \\ &= 1/2n\ [ab + a - b - (1)] \end{aligned} \tag{5.1}$$

***Main effect of B:***

$$\text{Effect of } B \text{ at the low level of } A = \frac{[b - (1)]}{n}$$

$$\text{Effect of } B \text{ at the high level of } A = \frac{(ab - a)}{n}$$

$$\begin{aligned} \text{Average effect of } B &= 1/2n\ [b - (1) + (ab - a)] \\ &= 1/2n\ [ab + b - a - (1)] \end{aligned} \tag{5.2}$$

***Interaction effect (AB):***

Interaction effect $AB$ is the average difference between the effect of $A$ at the high level of $B$ and the effect of $A$ at the low level of $B$.

$$AB = 1/2n\ [(ab - b) - \{a - (1)\}]$$
$$= 1/2n\ [ab + (1) - a - b] \qquad (5.3)$$

We can also obtain interaction effect $AB$ as the average difference between the effect of $B$ at the high level of $A$ and the effect of $B$ at the low level of $A$.

$$AB = 1/2n\ [(ab - a) - \{b - (1)\}]$$
$$= 1/2n\ [ab + (1) - a - b] \qquad (5.4)$$

which is same as Eq. (5.3).

Observe that the quantities in the parenthesis of Eqs. (5.1), (5.2) and (5.3) are the contrasts used to estimate the effects. Thus, the contrasts used to estimate the main and interaction effects are:

Contrast $A$ ($C_A$) = $ab + a - b - (1)$
Contrast $B$ ($C_B$) = $ab + b - a - (1)$
Contrast $AB$ ($C_{AB}$) = $ab + (1) - a - b$

Note that these three contrasts are also orthogonal.

We may now obtain the value of the contrasts by substituting the relevant data from Figure 5.1.

$C_A = 80 + 93 - 113 - 77 = -17$
$C_B = 80 + 113 - 93 - 77 = 23$
$C_{AB} = 80 + 77 - 93 - 113 = -49$

The average effects are estimated by using the following expression.

$$\text{Average effect of any factor} = \frac{\text{contrast}}{n\,2^{k-1}}$$

$$= \frac{\text{contrast}}{2n}, \text{ since } k = 2$$

which is same as Eqs. (5.1) or (5.2) or (5.3)

Therefore,

Effect of $A$ = –17/2 (2) = –4.25
Effect of $B$ = 23/4 = 5.75
Effect of $AB$ = –49/4 = –12.25

Observe that the effect of $A$ is negative, indicating that increasing the factor from low level (60°C) to high level (70°C) will decrease the yield. The effect of $B$ is positive suggesting that increasing $B$ from low level to high level will increase the yield. Compared to main effects the interaction effect is higher and negative. This indicates that the behaviour of response is different at the two levels of the other factor. That is, increasing $A$ from low level to high level results in increase of yield at low level of factor $B$ and decrease of yield at high level of factor $B$. With respect to factor $B$, the result is opposite.

In these $2^k$ designs, the magnitude as well as the direction of the factor effects indicate which variables are important. This can be confirmed by analysis of variance. As already discussed in Chapter 3, we can compute sum of squares for any effect ($SS_{\text{effect}}$) using the contrast.

$$SS_{\text{effect}} = \frac{(\text{contrast})^2}{n\Sigma c_i^2} \tag{5.5}$$

where $c_i$ is the coefficient of the $i$th treatment present in the contrast

In $2^k$ design,
$$SS_{\text{effect}} = \frac{(\text{contrast})^2}{n2^k} \tag{5.6}$$

where $n$ is the number of replications.

***Computation of sum of squares***

$$SS_A = \frac{(C_A)^2}{n2^k} = \frac{(-17)^2}{2 \times 4} = 36.125$$

$$SS_B = \frac{(C_B)^2}{n2^k} = \frac{(23)^2}{2 \times 4} = 66.125$$

$$SS_B = \frac{(C_{AB})^2}{n2^k} = \frac{(-49)^2}{2 \times 4} = 300.125$$

$$SS_{\text{Total}} = \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{n} Y_{ijk}^2 - \text{CF} \tag{5.7}$$

$$= (40)^2 + (37)^2 + (43)^2 + (50)^2 + \dots + (43)^2 - \frac{(363)^2}{8}$$

$$= 16933.000 - 16471.125$$

$$= 461.875$$

$$SS_e = SS_{\text{Total}} - SS_A - SS_B - SS_{AB}$$

$$= 461.875 - 36.125 - 66.125 - 300.125$$

$$= 59.500$$

The analysis of variance is summarized below (Table 5.1).

**TABLE 5.1** ANOVA for Illustration 5.1

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ |
|---|---|---|---|---|
| Temperature ($A$) | 36.125 | 1 | 36.125 | 2.43 |
| Pressure ($B$) | 66.125 | 1 | 66.125 | 4.45 |
| Interaction $AB$ | 300.125 | 1 | 300.125 | 20.18 |
| Error | 59.5 | 4 | 14.875 | |
| Total | 461.875 | 7 | | |

Since $F_{0.05,1,4} = 7.71$, only the interaction is significant and main effects are not significant.

### 5.2.2 Development of Contrast Coefficients

It is often convenient to write the treatment combinations in the order (1), *a*, *b*, *ab* referred to as standard order. Using this order, the contrast coefficients used in estimating the effects above are as follows:

| *Effect* | (1) | *a* | *b* | *ab* |
|---|---|---|---|---|
| *A* | –1 | +1 | –1 | +1 |
| *B* | –1 | –1 | +1 | +1 |
| *AB* | +1 | –1 | –1 | +1 |

Note that the contrast coefficients for *AB* are the product of the coefficients of *A* and *B*. The contrast coefficients are always either +1 or –1. Now, we can develop a table of plus and minus signs (Table 5.2) that can be used to find the contrasts for estimating any effect.

**TABLE 5.2** Algebraic signs to find contrasts

| *Treatment* | *Factorial effect* | | | |
|---|---|---|---|---|
| *combination* | *I* | *A* | *B* | *AB* |
| (1) | + | – | – | + |
| *a* | + | + | – | – |
| *b* | + | – | + | – |
| *ab* | + | + | + | + |

Note that the column headings *A*, *B* and *AB* are the main and interaction factorial effects. *I* stands for the identity column with all the +ve signs under it. The treatment combinations should be written in the standard order. To fill up the table, under effect *A* enter plus (+) sign against all treatment combinations containing *a*. Against other treatment combinations enter minus (–) sign. Similarly under *B*, enter plus (+) against treatment combinations containing *b* and minus (–) sign against others, signs under *AB* are obtained by multiplying the respective signs under *A* and *B*. To find the contrast for estimating any effect, simply multiply the signs in the appropriate column of Table 5.2 by the corresponding treatment combination and add. For example, to estimate *A*, the contrast is $-(1) + a - b + ab$ which is same as Eq. (5.1). Similarly, other contrasts can be obtained. From these contrasts sum of squares are computed as explained before.

### 5.2.3 The Regression Model

The regression approach is very useful when the factors are quantitative. The model can be used to predict response at the intermittent levels of the factors. For Illustration 5.1, the regression model is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + e \tag{5.8}$$

where $X_i$ = coded variable

$\beta s$ = regression coefficients

$X_1$ = factor $A$ (temperature) and
$X_2$ = factor $B$ (pressure)

The relationship between the natural variables (temperature and the pressure) and the coded variables is

$$X_1 = \frac{A - \dfrac{(A_H + A_L)}{2}}{\dfrac{(A_H - A_L)}{2}} \tag{5.9}$$

$$X_2 = \frac{B - \dfrac{(B_H + B_L)}{2}}{\dfrac{(B_H - B_L)}{2}} \tag{5.10}$$

where, $A$ = temperature and
$B$ = pressure.
$A_H$ and $A_L$ = high and low level values of factor $A$
$B_H$ and $B_L$ = high and low level values of factor $B$.

When the natural variables have only two levels, the levels of the coded variable will be ±1. The relationship between the natural and coded variables for Illustration 5.1 is

$$X_1 = \frac{\text{Temperature} - (70 + 60)/2}{(70 - 60)/2} = \frac{\text{Temperature} - 65}{5} \tag{5.11}$$

$$X_2 = \frac{\text{Pressure} - (150 + 100)/2}{(150 - 100)/2} = \frac{\text{Pressure} - 125}{25} \tag{5.12}$$

When temperature is at high level (70°C), $X_1 = +1$ and when temperature is at low level $X_1 = -1$. Similarly, $X_2$ will be +1 when pressure is at high level and –1 when pressure is at low level.

The complete regression model for predicting yield is

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 \tag{5.13}$$

$\beta_0$ is estimated by

$$\text{Grand average} = \frac{\text{Grand total}}{N} \tag{5.14}$$

Other coefficients are estimated as one-half of their respective effects.

$$\hat{\beta}_1 = \frac{\text{Effect of } A}{2} \tag{5.15}$$

$$\hat{\beta}_2 = \frac{\text{Effect of } B}{2} \tag{5.16}$$

$$\hat{\beta}_3 = \frac{\text{Effect of } AB}{2} \tag{5.17}$$

Therefore, $\hat{\beta}_0 = \dfrac{363}{8} = 45.375.$

$$\beta_1 = \frac{4.25}{2}, \ \beta_2 = \frac{5.75}{2}, \text{ and } \beta_3 = -\frac{12.25}{2}$$

Hence, the predictive model for Illustration 5.1 is

$$\hat{Y} = 45.375 - 2.125\ X_1 + 2.875\ X_2 - 6.125\ X_1X_2 \tag{5.18}$$

In terms of natural variables, the model is

$$\hat{Y} = 45.375 - 6.125 \left[\frac{(\text{Temperature} - 65)}{5}\right]\left[\frac{(\text{Pressure} - 125)}{25}\right] \tag{5.19}$$

Equation (5.19) can be used to construct the contour plot and response surface using any statistical software package. Note that only the significant effect is included in Eq. (5.19).

Equation (5.18) can be used to compute the residuals. For example, when both temperature and pressure are at low level, $X_1 = -1$ and $X_2 = -1$;

$$\hat{Y} = 45.375 - 2.125\ (-1) + 2.875\ (-1) - 6.125\ (-1)\ (-1) = 38.5$$

There are two observations of this treatment combination and hence the two residuals are

$$R_1 = 40 - 38.5 = 1.5$$

$$R_2 = 37 - 38.5 = -1.5$$

Similarly, the residuals at other treatment combinations can be obtained.

$X_1 = +1$ and $X_2 = -1$, the observations are 43 and 50 respectively

$$\hat{Y} = 46.5$$

$$R_3 = 43 - 46.5 = -3.5$$

$$R_4 = 50 - 46.5 = 3.5$$

When $X_1 = -1$ and $X_2 = +1$,

$$\hat{Y} = 56.5$$

$$R_5 = 59 - 56.5 = 2.5$$

$$R_6 = 54 - 56.5 = -2.5$$

At $X_1 = +1$ and $X_2 = +1$,

$$\hat{Y} = 40$$

$$R_7 = 37 - 40 = -3.0$$

$$R_8 = 43 - 40 = 3.0$$

These residuals can be analysed as discussed earlier in Chapter 3. Figures 5.2 to 5.5 show the normal probability plot of residuals, plot of residuals vs fitted value, main effects plot and interaction effect plot respectively for Illustration 5.1.

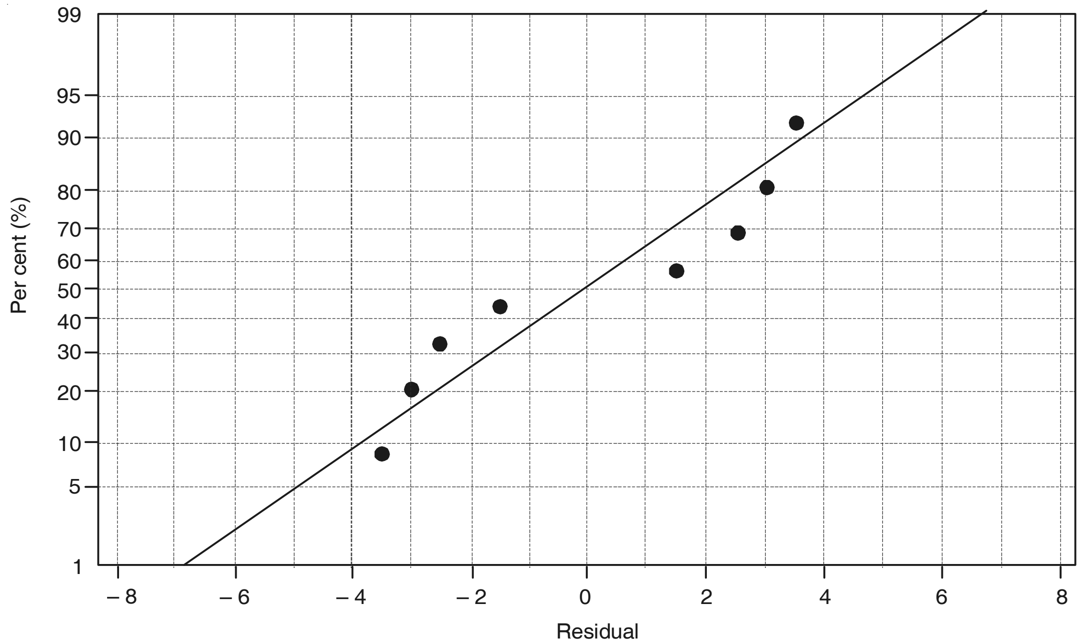



**FIGURE 5.2** Normal probability plot of residuals for Illustration 5.1.

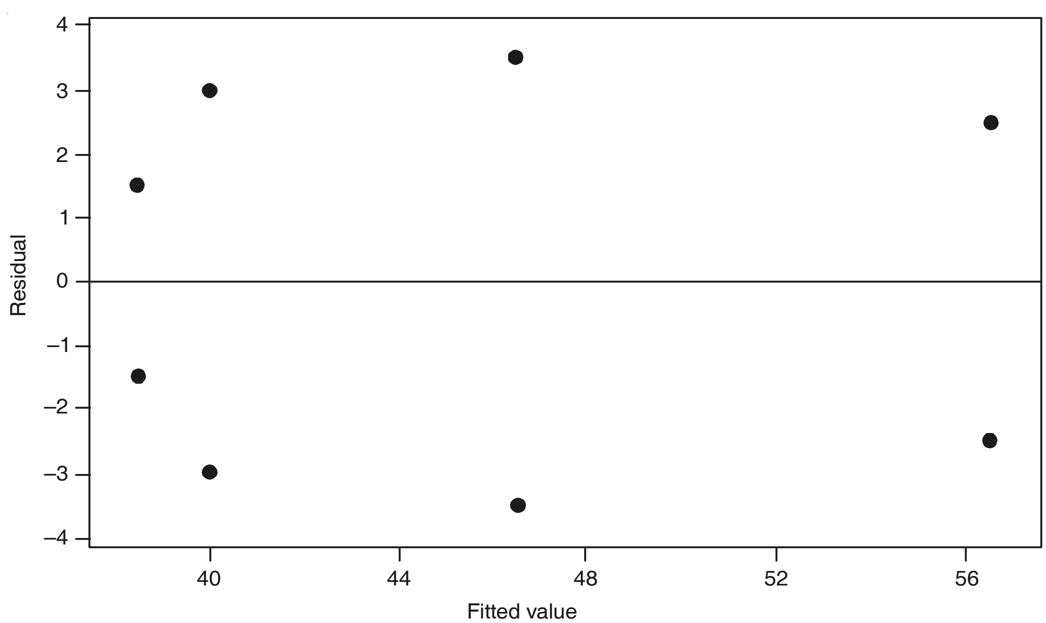



**FIGURE 5.3** Plot of residuals vs. fitted values for Illustration 5.1.

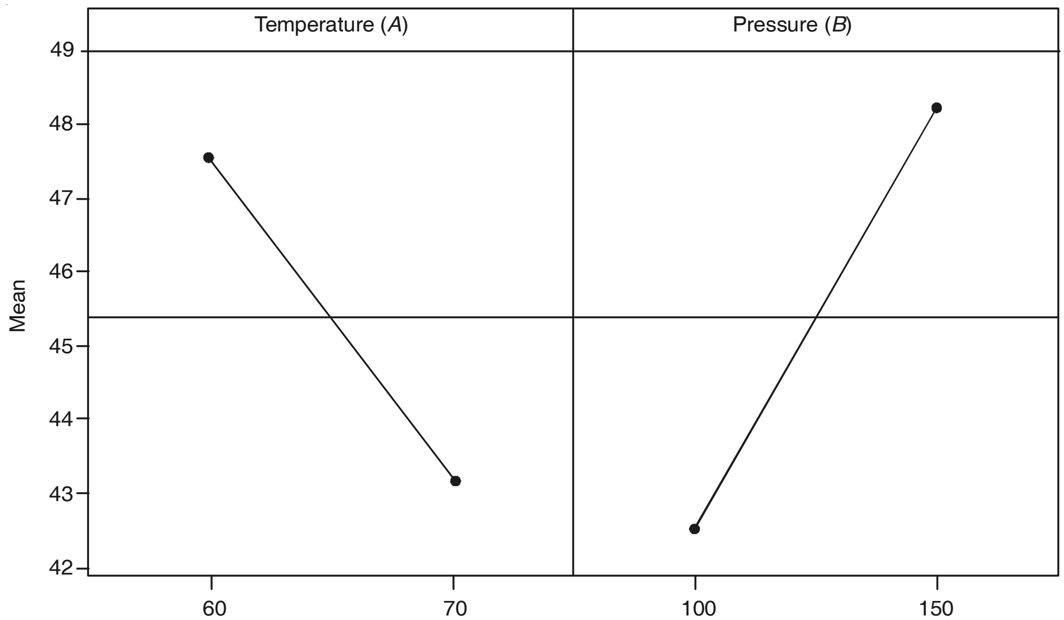



**FIGURE 5.4** Plot of main effects for Illustration 5.1.

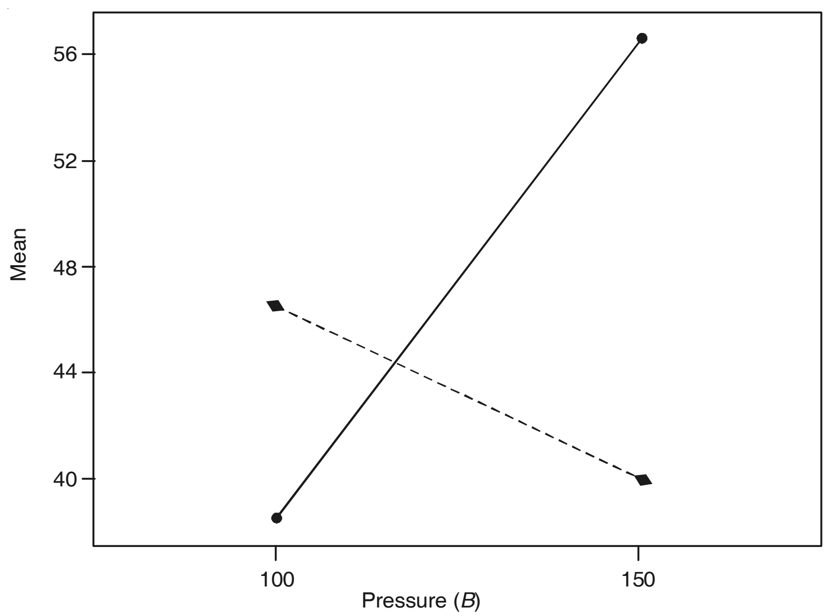



**FIGURE 5.5** Plot of Interaction effect for Illustration 5.1.

Equation (5.19) can be used to construct the contour plot and response surface using any computer software package. Figure 5.6 shows the contour plot of the response and Figure 5.7 shows the surface plot of response for Illustration 5.1. Since interaction is significant, we have

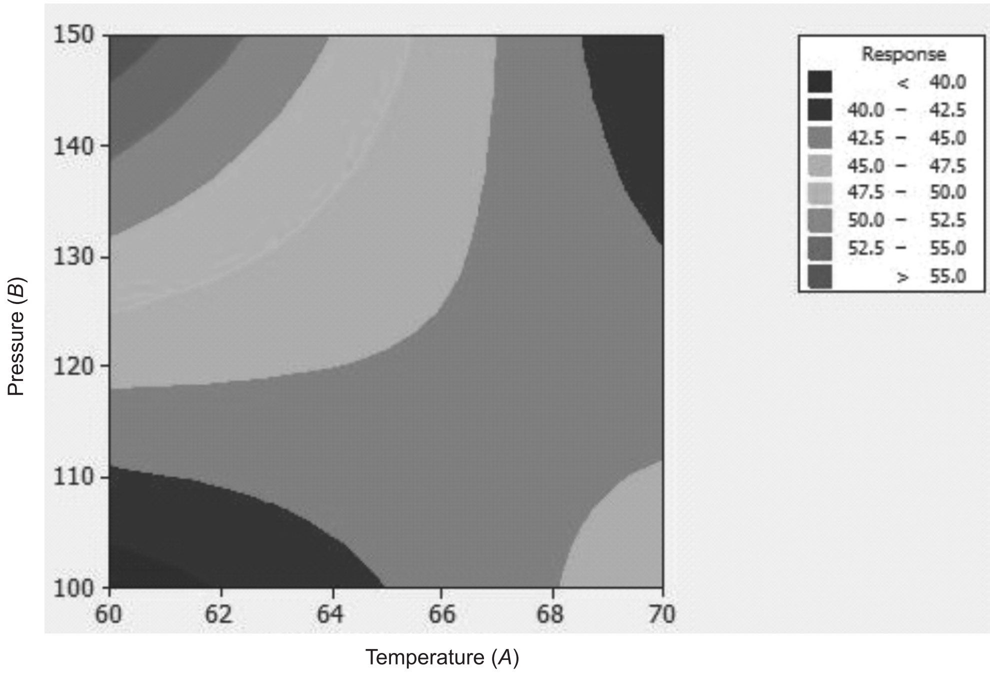



**FIGURE 5.6** Contour plot of response for Illustration 5.1.

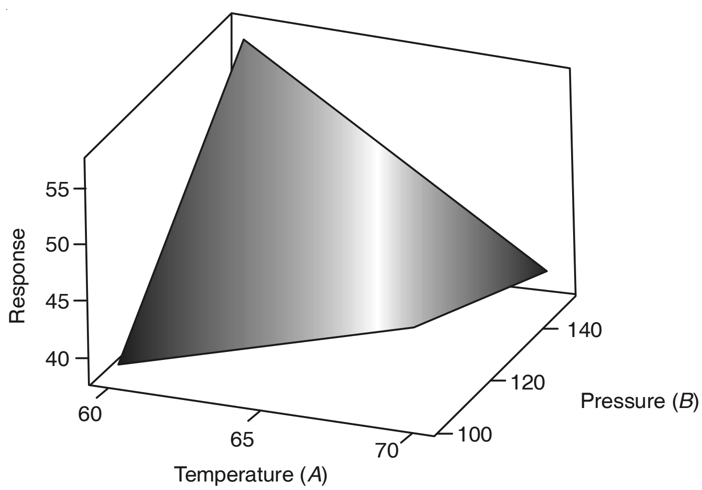



**FIGURE 5.7** Surface plot of response for Illustration 5.1.

curved contour lines of constant response and twisted three dimensional response surface. If the interaction is not significant, the response surface will be a plane and the contour plot contain parallel straight lines.

## 5.3 THE $2^3$ FACTORIAL DESIGN

Suppose we have three factors $A$, $B$ and $C$, each at two levels. This design is called $2^3$ *factorial design*. In this case we have $2^3 = 8$ treatment combinations, shown below in the standard order.

$$(1), a, b, ab, c, ac, bc, abc$$

For analysing the data from $2^k$ design up to $k \leq 4$, usually we follow the procedure discussed in $2^2$ designs. The analysis of $2^3$ designs is explained in the following sections. If $k > 4$, use of computer software is preferable.

### 5.3.1 Development of Contrast Coefficient Table

We can construct a plus–minus table (Table 5.3) by which the contrasts can be formed for further analysis as in the case of $2^2$ designs.

**TABLE 5.3** Algebraic signs to find contrasts for a $2^3$ design

| *Treatment* | *Factorial effect* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *combination* | $I$ | $A$ | $B$ | $AB$ | $C$ | $AC$ | $BC$ | $ABC$ |
| (1) | + | – | – | + | – | + | + | – |
| $a$ | + | + | – | – | – | – | + | + |
| $b$ | + | – | + | – | – | + | – | + |
| $ab$ | + | + | + | + | – | – | – | – |
| $c$ | + | – | – | + | + | – | – | + |
| $ac$ | + | + | – | – | + | + | – | – |
| $bc$ | + | – | + | – | + | – | + | – |
| $abc$ | + | + | + | + | + | + | + | + |

Table 5.3 has the following properties:

- Every column has got equal number of plus and minus signs except for column $I$.
- Sum of product of signs in any two columns is zero.
- Column $I$ is multiplied by any other column leaves that column unchanged. That is, $I$ is an identity element.
- The product of any two columns modulus 2 yields a column in the table.

  For example, $B * C = BC$ and $BC * B = B^2 C = C$

All these properties indicate the orthogonality of the contrasts used to estimate the effects.

$$\text{Average effect of any factor or effect} = \frac{\text{contrast}}{(n * 2^{k-1})} = \frac{C}{(4 * n)} \tag{5.20}$$

$$\text{Sum of squares of any factor/effect} = \frac{(\text{contrast})^2}{n * 2^k} = \frac{C^2}{(8 * n)} \tag{5.21}$$

where $n$ is the number of replications.

## ILLUSTRATION 5.2

**A $2^3$ Design: Surface Roughness Experiment**

An experiment was conducted to study the effect of tool type ($A$), speed ($B$) and feed ($C$) on the surface finish of a machined part. Each factor was studied at two levels and obtained two replications. The order of experiment was random. The results obtained are given in Table 5.4.

**TABLE 5.4** Surface roughness measurements

| | *Tool type* ($A$) | | | |
|---|---|---|---|---|
| | 1 | | 2 | |
| *Feed* (mm/rev) ($C$) | *Speed* (rpm) ($B$) | | *Speed* (rpm) ($B$) | |
| | 1000 | 1200 | 1000 | 1200 |
| 0.10 | 54 | 41 | 60 | 43 |
| | 73 | 51 | 53 | 49 |
| | (1) = 127 | $b$ = 92 | $a$ = 113 | $ab$ = 92 |
| 0.20 | 86 | 63 | 82 | 66 |
| | 66 | 65 | 73 | 65 |
| | $c$ = 152 | $bc$ = 128 | $ac$ = 155 | $abc$ = 131 |

Table 5.4 also gives the respective treatment totals. The response total for each treatment combination for Illustration 5.2 is summarized in Table 5.5.

**TABLE 5.5** Response totals for Illustration 5.2

| Treatment combination | (1) | $a$ | $b$ | $ab$ | $c$ | $ac$ | $bc$ | $abc$ |
|---|---|---|---|---|---|---|---|---|
| Response total | 127 | 113 | 92 | 92 | 152 | 155 | 128 | 131 |

To analyse the data using the ANOVA, the method of contrasts is followed. The contrasts are obtained from the plus–minus table (Table 5.3).

***Evaluation of contrasts***

$$C_A = -(1) + a - b + ab - c + ac - bc + abc$$
$$= -127 + 113 - 92 + 92 - 152 + 155 - 128 + 131 = -8$$
$$C_B = -(1) - a + b + ab - c - ac + bc + abc$$
$$= -127 - 113 + 92 + 92 - 152 - 155 + 128 + 131 = -104$$
$$C_C = -(1) - a - b - ab + c + ac + bc + abc$$
$$= -127 - 113 - 92 - 92 + 152 + 155 + 128 + 131 = 142$$
$$C_{AB} = +(1) - a - b + ab + c - ac - bc + abc$$
$$= 127 - 113 - 92 + 92 + 152 - 155 - 128 + 131 = 14$$
$$C_{AC} = +(1) - a + b - ab - c + ac - bc + abc$$
$$= 127 - 113 + 92 - 92 - 152 + 155 - 128 + 131 = 20$$
$$C_{BC} = +(1) + a - b - ab - c - ac + bc + abc$$
$$= 127 + 113 - 92 - 92 - 152 - 155 + 128 + 131 = 8$$

$$C_{ABC} = -(1) + a + b - ab + c - ac - bc + abc$$
$$= -127 + 113 + 92 - 92 + 152 - 155 - 128 + 131 = -14$$

***Computation of effects***

$$\text{Factor effect} = \frac{\text{contrast}}{(n * 2^{k-1})} = \frac{\text{contrast}}{(2 * 4)}$$

$A = -8/8 = -1.00$, $B = -104/8 = -13$, $C = 142/8 = 17.75$,
$AB = 14/8 = 1.75$, $AC = 20/8 = 2.5$, $BC = 8/8 = 1.00$,
$ABC = -14/8 = -1.75$

***Computation of sum of squares***

$$\text{Factor sum of squares} = \frac{(\text{contrast})^2}{(n * 2^k)} = \frac{(C)^2}{(2 * 8)}$$

$SS_A = -8^2/16 = 4.00$, $SS_B = (-104)^2/16 = 676.00$, $SS_c = (142)^2/16 = 1260.25$,
$SS_{AB} = (14)^2/16 = 12.25$, $SS_{AC} = (20)^2/16 = 25.00$, $SS_{BC} = (8)^2/16 = 4.00$,
$SS_{ABC} = (-14)^2/16 = 12.25$

The total sum of squares is computed as usual.

$$SS_{\text{Total}} = \frac{(54)^2 + (73)^2 + (41)^2 + \cdots + (66)^2 + (65)^2 - (990)^2}{16}$$
$$= 63766 - 61256.25$$
$$= 2509.75$$

$$SS_e = SS_{\text{Total}} - (SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC} - SS_{ABC})$$
$$= 2509.75 - 1993.75$$
$$= 516.00$$

The computations are summarized in the ANOVA Table 5.6

**TABLE 5.6** Analysis of variance for Illustration 5.2 ($2^3$ Design)

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ | *Contribution*** (%) |
|---|---|---|---|---|---|
| Type of tool ($A$) | 4.00 | 1 | 4.00 | 0.06* | 0.16 |
| Speed ($B$) | 676.00 | 1 | 676.00 | 10.48 | 26.93 |
| Feed ($C$) | 1260.25 | 1 | 1260.25 | 19.54 | 50.21 |
| $AB$ | 12.25 | 1 | 12.25 | 0.19* | 0.49 |
| $AC$ | 25.00 | 1 | 25.00 | 0.39* | 1.00 |
| $BC$ | 4.00 | 1 | 4.00 | 0.06* | 0.16 |
| $ABC$ | 12.25 | 1 | 12.25 | 0.19* | 0.49 |
| Pure error | 516.00 | 8 | 64.50 | — | 20.56 |
| Total | 2509.75 | 15 | | | 100.00 |

* Insignificant ** Contribution = $\dfrac{\text{Factor sum of squares}}{\text{Total sum of squares}} \times 100$

At 5% significance level, only the speed (*B*) and feed (*C*) significantly influence the surface finish. Type of tool and all interactions have no effect on surface finish.

The column Contribution in the ANOVA table indicates the percentage contribution of each term in the model to the total sum of squares. Percentage contribution is an appropriate measure of the relative importance of each term in the model. In Illustration 5.2 note that the factors speed (*B*) and feed (*C*) together account for about 77% of the total variability. Also note that the error contributes to about 20%. According to experts in this field, if the error contribution is more than 15%, we can suspect that some factors which might have influence on the response are not studied. And thus there will be future scope for improving the results. If the error contribution is significantly less, the scope for further improvement of the performance of the process or product would be marginal.

### 5.3.2 Yates Algorithm for the $2^k$ Design

Yates (1937) has developed an algorithm for estimating the factor effects and sum of squares in a $2^k$ design. Let us apply this algorithm to the $2^3$ design illustration on surface roughness experiment (Table 5.4).

***Procedure:*** The following steps discuss Yates algorithm:

*Step* 1: Create a table (Table 5.7) with the first column containing the treatment combinations arranged in the standard order.
*Step* 2: Fill up the second column of the table with the response total (sum of replications of each experiment if more than one replication is taken)
*Step* 3: Label the next $k$ columns as 1, 2, …, $k$ equal to the number of factors studied in the experiment.
*Step* 4: Label the column that follow the $k$th column as an effect column (indicates the effects)
*Step* 5: The last but one column gives the effect estimate
*Step* 6: The last column gives the sum of squares

Table 5.7 gives the implementation of Yates algorithm for Illustration 5.2.

**TABLE 5.7** Yates algorithm for the surface roughness experiment

| *Treatment combinations* | *Response total* | (1) | (2) | (3) | *Effect* | *Estimate of effect* $(3) \div n2^{k-1}$ | *Sum of squares* $(3)^2 \div n2^k$ |
|---|---|---|---|---|---|---|---|
| (1) | 127 | 240 | 424 | 990 | *I* | — | — |
| *a* | 113 | 184 | 566 | –8 | *A* | –1.00 | 4.00 |
| *b* | 92 | 307 | –14 | – 104 | *B* | –13.00 | 676.00 |
| *ab* | 92 | 259 | 6 | 14 | *AB* | 1.75 | 12.25 |
| *c* | 152 | –14 | –56 | 142 | *C* | 17.75 | 1260.25 |
| *ac* | 155 | 0 | – 48 | 20 | *AC* | 2.50 | 25.00 |
| *bc* | 128 | 3 | 14 | 8 | *BC* | 1.00 | 4.00 |
| *abc* | 131 | 3 | 0 | –14 | *ABC* | –1.75 | 12.25 |

The following are the explanation for the entries in Table 5.7:

*Treatment combinations*: Write in the standard order.

*Response*: Enter the corresponding treatment combination totals (if one replication, that value is entered)

*Column* (1): Entries in upper half is obtained by adding the responses of adjacent pairs.

(127 + 113 = 240), (92 + 92 = 184), etc.

Entries in lower half is obtained by changing the sign of the first entry in each of the pairs in the response and add the adjacent pairs.

(–127 + 113 = –14), (–92 + 92 = 0) etc.

*Column* (2): Use values in column (1) and obtain values in column (2) similar to column (1)

*Column* (3): Use values in column (2) and obtain column (3) following the same procedure as in Column (1) or Column (2).

Note that the first entry in column (3) is the grand total and all others are the corresponding contrast totals. Note that the contrast totals, effect estimates and the sum of squares given in Table 5.7 are same as those obtained earlier in Illustration 5.2. This algorithm is applicable only to $2^k$ designs.

### 5.3.3 The Regression Model

The full (complete) regression model for predicting the surface finish is given by

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2 + \beta_5 X_1 X_3 + \beta_6 X_2 X_3 + \beta_7 X_1 X_2 X_3 \tag{5.22}$$

where $\beta_0$ is overall mean.

All other coefficients ($\beta$s) are estimated as one half of the respective effects as explained earlier.

If all terms in the model are used (significant and non-significant) for predicting the response, it may lead to either over estimation or under estimation. Hence, for predicting only the significant effects are considered. So the model for predicting the surface finish is

$$\hat{Y} = \beta_0 + \beta_2 X_2 + \beta_3 X_3 \tag{5.23}$$

$$= 61.87 + \left(\frac{-13}{2}\right) X_2 + \left(\frac{17.75}{2}\right) X_3$$

where the coded variables $X_2$ and $X_3$ represent the factors $B$ and $C$ respectively.

Residuals can be obtained as the difference between the observed and predicted values of surface roughness. The residuals are then analysed as usual.

## 5.4 STATISTICAL ANALYSIS OF THE MODEL

$SS_{\text{model}}$ is the summation of all the treatment sum of squares which is given as:

$$SS_{\text{model}} = SS_A + SS_B + SS_C + SS_{AB} + SS_{AC} + SS_{BC} + SS_{ABC}$$
$$= 1993.75$$

$$F_0 = \frac{MS_{\text{model}}}{MS_e} = \frac{SS_{\text{model}}/df_{\text{model}}}{SS_{\text{PE}}/df_{\text{PE}}} \tag{5.24}$$

where PE is the pure error (error due to repetitions, also called experimental error) is used to test the hypothesis.

$$F_0 = \frac{1993.75/7}{516.00/8}$$

$$= \frac{284.82}{64.50} = 4.42$$

$F_{0.05,7,8} = 3.50$. Since $F_0$ is greater than $F_{0.05,7,8}$ (5% significance level), the model is significant. The hypotheses for testing the regression coefficients are

$$H_0\text{: } \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$$

$$H_1\text{: at least one } \beta \neq 0.$$

The revised ANOVA with the significant effects is given in Table 5.8. Note that the sum of squares of the insignificant effects are pooled together as lack of fit and added to the pure sum of squares of error and shown as pooled error sum of squares. The corresponding degrees of freedom are also pooled and added. This pooled error sum of squares is used to test the effects.

**TABLE 5.8** Revised ANOVA for Illustration 5.2

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ | $\alpha = 5\%$ |
|---|---|---|---|---|---|
| Speed (*B*) | 676.00 | 1 | 676.00 | 15.32 | Significant |
| Feed (*C*) | 1260.25 | 1 | 1260.25 | 28.58 | Significant |
| Pooled error | 573.50 | 13 | 44.12 | | |
| Lack of fit | 57.50 | 5 | 11.50 | 0.18 | Not significant |
| Pure error | 516.00 | 8 | 64.50 | | |
| Total | 2509.75 | 15 | | | |

Note that the effects are tested with pooled error mean square and the lack of fit is tested with pure error mean square.

***Other statistics***

- Coefficient of determination ($R^2$):

$$R^2 \text{ (full model)} = SS_{\text{model}}/SS_{\text{total}} \tag{5.25}$$

$$= \frac{1993.75}{2509.75} = 0.7944$$

$R^2$ is a measure of the proportion of total variability explained by the model. The problem with this statistic is that it always increases as the factors are added to the model even if these are insignificant factors. Therefore, another measure $R^2_{\text{adj}}$(adjusted $R^2$) is recommended.

$$R^2_{\text{adj}} = \frac{1 - SS_{\text{PE}}/df_{\text{PE}}}{SS_{\text{Total}}/df_{\text{Total}}} \tag{5.26}$$

$$= \frac{(1 - 516/8)}{(2509.75/15)}$$

$$= \frac{(1 - 64.50)}{167.32} = 0.6145$$

$R^2_{\text{adj}}$ is a statistic that is adjusted for the size of the model, that is, the number of factors. Its value decreases if non-significant terms are added to a model.

- *Prediction Error Sum of Squares* (*PRESS*): It is a measure of how well the model will predict the new data. It is computed from the prediction errors obtained by predicting the *i*th data point with a model that includes all observations except the *i*th one. A model with a less value of PRESS indicates that the model is likely to be a good predictor. The prediction $R^2$ statistic is computed using Eq. (5.27).

$$R^2_{\text{Pred}} = \frac{1 - \text{PRESS}}{SS_{\text{Total}}} \tag{5.27}$$

Usually, computer software is used to analyse the data which will generate all the statistics.

The software package will also give the standard error ($S_e$) of each regression coefficient and the confidence interval.

The standard error of each coefficient, $S_e(\hat{\beta})$ is computed from Eq. (5.28).

$$S_e(\hat{\beta}) = \sqrt{V(\hat{\beta})} = \sqrt{\frac{MS_e}{n2^k}} \tag{5.28}$$

where, $MS_e$ = pure error mean square and
$n$ = number of replications

The 95% confidence interval on each regression coefficient is computed from Eq. (5.29).

$$-\hat{\beta} - t_{0.025, N-P}\, S_e(\hat{\beta}) \le \beta \le \hat{\beta} + t_{0.025,\, N-P}\, S_e(\hat{\beta}) \tag{5.29}$$

where $N$ = total number of runs in the experiment, 16 in this case and
$P$ = number of model parameters including $\beta_0$ which equal to 8.

For Illustration 5.2

$$S_e(\hat{\beta}) = \sqrt{\frac{64.50}{2 \times 2^3}} = 2.01$$

$N$ = 16 (including replications) and $P$ = 8.

Substituting these values, we can obtain confidence intervals for all the coefficients.

## 5.5 THE GENERAL $2^k$ DESIGN

The analysis of $2^3$ design can be extended to the case of a $2^k$ factorial design ($k$ factors, each at two levels)

The statistical model for a $2^k$ design will have

- $k$ main effects
- $\binom{K}{2}$ is the number of combinations of $K$ items taken 2 at a time = two-factor interactions
- $\binom{K}{3}$ three-factor interactions … and one $k$-factor interaction.

***Procedure***

1. Estimate the factor effects and examine their signs and magnitude. This gives an idea about the important factors and interactions and in which direction these factors to be adjusted to improve response.
2. Include all main effects and their interactions and perform ANOVA (replication of at least one design point is required to obtain error for testing).
3. Usually we remove all the terms which are not significant from the full model.
4. Perform residual analysis to check the adequacy of the model.
5. Interpret the results along with contour and response surface plots.

Usually we use statistical software for analyzing the model.

***Forming the contrasts:*** In $2^k$ design with $k \leq 4$, it is easier to form the plus–minus table to determine the required contrasts for estimating effects and sum of squares. For large values of $k$, forming plus–minus table is cumbersome and time consuming. In general, we can determine the contrast for an effect $AB \ldots k$ by expanding the right-hand side of Eq. (5.30),

$$\text{Contrast}_{AB \ldots K} = (a \pm 1)\ (b \pm 1)\ \ldots\ (k \pm 1) \tag{5.30}$$

Using ordinary algebra, Eq. (5.30) is expanded with 1 being replaced by (1) in the final expression, where (1) indicates that all the factors are at low level.

The sign in each set of parameters is negative if the factor is included in the effect and positive for the factors not included.

For example, in a $2^3$ design, the contrast $AC$ would be

$$\begin{aligned} \text{Contrast}_{AC} &= (a - 1)\ (c - 1)\ (b + 1) \\ &= abc + ac + b + (1) - ab - bc - a - c \end{aligned} \tag{5.31}$$

As another example, in a $2^4$ design, the contrast $ABD$ would be

$$\begin{aligned} \text{Contrast}_{ABD} &= (a - 1)\ (b - 1)\ (d - 1)\ (c + 1) \\ &= -(1) + a + b - ab - c + ac + bc - abc + d - ad - bd + abd + cd \\ &\quad - acd - bcd + abcd \end{aligned} \tag{5.32}$$

Once the contrasts are formed, we can estimate the effects and sum of squares as discussed earlier in this chapter.

## 5.6 THE SINGLE REPLICATE OF THE $2^k$ DESIGN

When the number of factors increases, the number of experimental runs would be very large. Especially in the screening experiments conducted to develop new processes or new products, the number of factors considered will usually be large since the experimenter has no prior knowledge on the problem. Because resources are usually limited, the experimenter may restrict the number of replications. When only one replication is obtained, we will not have the experimental error (pure error) for testing the effects.

A single replicate of a $2^k$ design is also called an *unreplicated factorial*. One approach to analyse the unreplicated factorial is to assume that certain higher order interactions are negligible and pool their sum of square to estimate the error. Occasionally, some higher order interactions may influence the response. In such cases, pooling higher order interactions is inappropriate. To overcome this problem, a different method has been suggested to identify the significant effects. It is suggested to obtain the normal probability plot of the effect estimates. Those effects which fall away from the straight line are considered as significant effects. Other effects are combined to estimate the error. Alternately, the half normal plot of effects can also be used in the place of normal probability plot. This is a plot of the absolute value of the effect estimated against their cumulative normal probabilities. The straight line on half normal plot always passes through the origin and close to the fiftieth percentile data value. Many analysts feel that the half normal plot is easier to interpret especially when there are only a few effect estimates.

**ILLUSTRATION 5.3**

**The $2^4$ Design with Single Replicate**

An experiment was conducted to study the effect of preheating ($A$), hardening temperature ($B$), quenching media ($C$) and quenching time ($D$) on the distortion produced in a gearwheel. The distortion was measured as the difference in the diameter before and after heat treatment (mm). Each factor was studied at two levels and only one replication was obtained. The results are given in Table 5.9.

**TABLE 5.9** Distortion data for Illustration 5.3

| *Treatment combination* | *Response (distortion)* | *Treatment combination* | *Response (distortion)* |
|---|---|---|---|
| (1) | 4.6 | *d* | 2.8 |
| *a* | 3.0 | *ad* | 3.8 |
| *b* | 5.6 | *bd* | 5.2 |
| *ab* | 5.6 | *abd* | 4.0 |
| *c* | 6.0 | *cd* | 4.3 |
| *ac* | 4.2 | *acd* | 3.8 |
| *bc* | 6.0 | *bcd* | 2.2 |
| *abc* | 6.0 | *abcd* | 4.0 |

We begin the analysis by computing factor effects and sum of squares. Let us use Yates algorithm for this purpose. Table 5.10 gives the computation of factor effects and sum of squares using Yates algorithm.

**TABLE 5.10** Yates algorithm for the $2^4$ design for Illustration 5.3

| *Treatment combinations* | *Response* | (1) | (2) | (3) | (4) | *Effect* | *Estimate of Effect* $(4) \div n2^{k-1}$ | *Sum of squares* $(4)^2 \div n2^k$ |
|---|---|---|---|---|---|---|---|---|
| (1) | 4.6 | 7.6 | 18.8 | 41.0 | 71.1 | *I* | — | — |
| *a* | 3.0 | 11.2 | 22.2 | 30.1 | –2.3 | *A* | –0.29 | 0.33 |
| *b* | 5.6 | 10.2 | 15.8 | –3.4 | 6.1 | *B* | 0.76 | 2.32 |
| *ab* | 5.6 | 12.0 | 14.3 | 1.1 | 3.5 | *AB* | 0.44 | 0.77 |
| *c* | 6.0 | 6.6 | –1.6 | 5.4 | 1.9 | *C* | 0.24 | 0.23 |
| *ac* | 4.2 | 9.2 | –1.8 | 0.7 | 1.3 | *AC* | 0.16 | 0.10 |
| *bc* | 6.0 | 8.1 | –0.2 | 3.4 | – 6.3 | *BC* | –0.79 | 2.48 |
| *abc* | 6.0 | 6.2 | 1.3 | 0.1 | 4.7 | *ABC* | 0.59 | 1.38 |
| *d* | 2.8 | –1.6 | 3.6 | 3.4 | –10.9 | *D* | –1.36 | 7.43 |
| *ad* | 3.8 | 0.0 | 1.8 | –1.5 | 4.5 | *AD* | 0.56 | 1.27 |
| *bd* | 5.2 | –1.8 | 2.6 | –0.2 | – 4.7 | *BD* | –0.59 | 1.38 |
| *abd* | 4.0 | 0.0 | –1.9 | 1.5 | –3.3 | *ABD* | –0.41 | 0.68 |
| *cd* | 4.3 | 1.0 | 1.6 | –1.8 | – 4.9 | *CD* | –0.61 | 1.50 |
| *acd* | 3.8 | –1.2 | 1.8 | – 4.5 | 1.7 | *ACD* | 0.21 | 0.18 |
| *bcd* | 2.2 | –0.5 | –2.2 | 0.2 | –2.7 | *BCD* | –0.34 | 0.46 |
| *abcd* | 4.0 | 1.8 | 2.3 | 4.5 | 4.3 | *ABCD* | 0.54 | 1.15 |

Since only one replication is taken, we cannot have experimental error for testing the effects. One way to resolve this issue is to pool the sum of squares of factors with less contribution into the error term as pooled error and test the other effects. This can be verified by plotting the effects on normal probability paper. Those effects which fall along the straight line on the plot are considered as insignificant and they can be combined to obtain pooled error. Effects which fall away from the straight line are considered as significant. In place of normal probability plot, half normal plot of effects can also be used. The computations required for normal probability plot is given in Table 5.11.

Figure 5.8 shows the normal probability plot of the effects. When the normal probability is plotted on a half-normal paper, we get half-normal probability plot. Figure 5.9 shows the half-normal plot of the effects for Illustration 5.3.

From the normal plot of the effects (Figure 5.8), it is observed that the effects *B*, *ABC* and *BC* fall away from the line. However, the effect with the largest magnitude (*D*) falls on the line indicating the draw back of using normal plot in identifying the significant effects. The half-normal plot of effects (Figure 5.9) clearly shows that the largest effect (*D*) is far away from all other effects indicating its significance. And we can suspect that the effects *BC* and *B* also as significant from Figure 5.9. However, this can be verified from ANOVA. Table 5.12 gives the initial ANOVA before pooling along with contribution.

From Table 5.12, it can be seen that the effects *B*, *BC* and *D* together contribute to about 57% of the total variation. Also these are considered as significant on the half-normal plot. Now, leaving these three effects (*D*, *BC* and *B*), all the other effects are pooled into the error term to test the effects. Table 5.13 gives ANOVA with the pooling error.

**TABLE 5.11** Normal probability computations of the effects for Illustration 5.3

| *J* | *Effect* | *Estimate of effect* | *Normal probability (%)* $(J - 0.5)/N$ |
|---|---|---|---|
| 1 | *D* | –1.36 | 3.33 |
| 2 | *BC* | –0.79 | 10.00 |
| 3 | *CD* | –0.61 | 16.67 |
| 4 | *BD* | –0.59 | 23.33 |
| 5 | *ABD* | –0.41 | 30.00 |
| 6 | *BCD* | –0.34 | 36.67 |
| 7 | *A* | –0.29 | 43.33 |
| 8 | *AC* | 0.16 | 50.00 |
| 9 | *ACD* | 0.21 | 56.67 |
| 10 | *C* | 0.24 | 63.33 |
| 11 | *AB* | 0.44 | 70.00 |
| 12 | *ABCD* | 0.54 | 76.67 |
| 1 | *AD* | 0.56 | 83.33 |
| 14 | *ABC* | 0.59 | 90.00 |
| 15 | *B* | 0.76 | 96.67 |

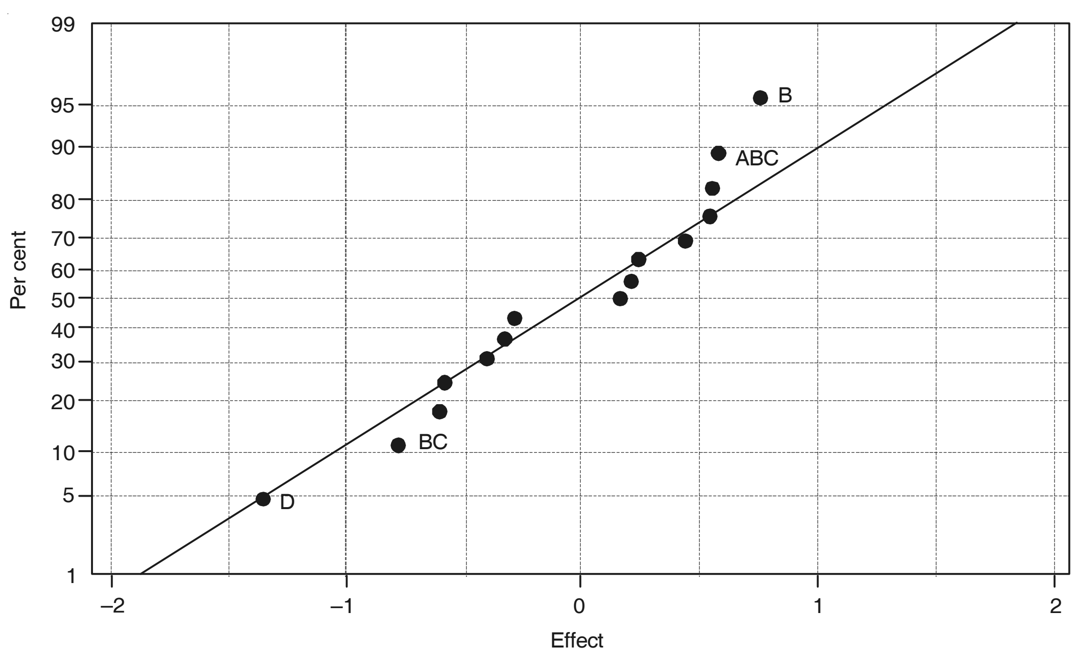



**FIGURE 5.8** Normal probability plot of effects for Illustration 5.3.

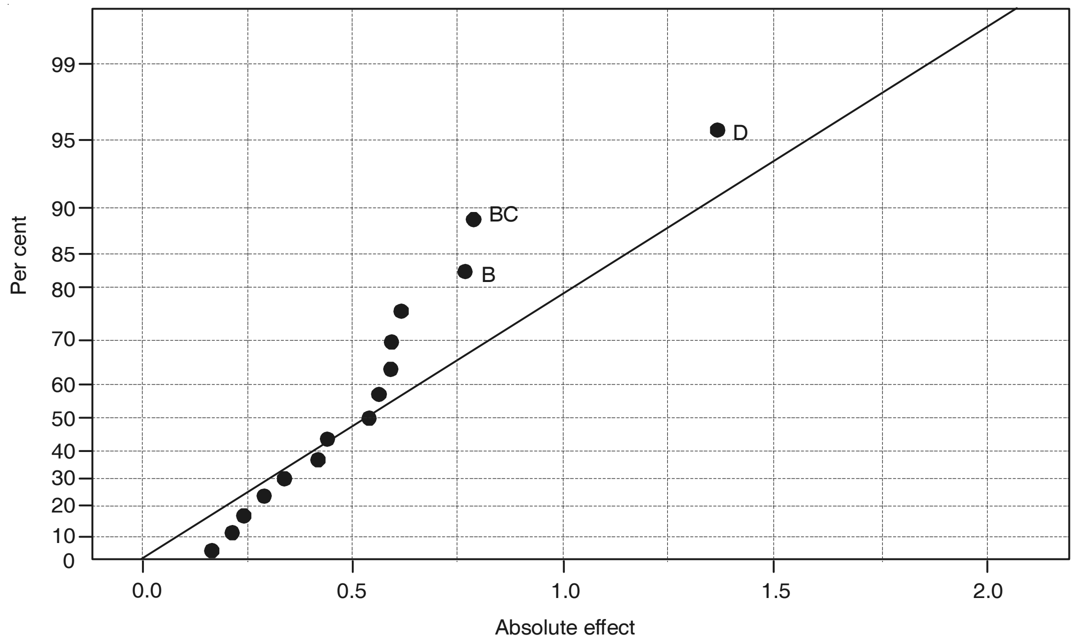



**FIGURE 5.9** Half-normal plot of effects for Illustration 5.3.

**TABLE 5.12** Analysis of variance for Illustration 5.3 ($2^4$ Design)

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Contribution** (%) |
|---|---|---|---|
| *A* | 0.33 | 1 | 1.52 |
| *B* | 2.32 | 1 | 10.71 |
| *AB* | 0.77 | 1 | 3.55 |
| *C* | 0.23 | 1 | 1.06 |
| *AC* | 0.10 | 1 | 0.46 |
| *BC* | 2.48 | 1 | 11.45 |
| *ABC* | 1.38 | 1 | 6.37 |
| *D* | 7.43 | 1 | 34.30 |
| *AD* | 1.27 | 1 | 5.86 |
| *BD* | 1.38 | 1 | 6.37 |
| *ABD* | 0.68 | 1 | 3.14 |
| *CD* | 1.50 | 1 | 6.93 |
| *ACD* | 0.18 | 1 | 0.83 |
| *BCD* | 0.46 | 1 | 2.12 |
| *ABCD* | 1.15 | 1 | 5.30 |
| Total | 21.66 | | |

**TABLE 5.13** Analysis of variance after pooling for Illustration 5.3

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ |
|---|---|---|---|---|
| $B$ | 2.33 | 1 | 2.33 | 2.96 |
| $BC$ | 2.48 | 1 | 2.48 | 3.15 |
| $D$ | 7.42 | 1 | 7.42 | 9.44 |
| Pooled error | 9.43 | 12 | 0.786 | |

At 5% level of significance ($F_{1,12} = 4.75$), only factor $D$ is significant as revealed by the half-normal plot. Our initial claim about $B$ and $BC$ is turned out to be false. So, half-normal plot can be a reliable tool for identifying true significant effects.

## 5.7 ADDITION OF CENTER POINTS TO THE $2^K$ DESIGN

Usually, we assume linearity in factor effects of $2^k$ design. When factors are quantitative, we fit a first order model

$$Y = \beta_0 + \sum_{j=1}^{K} \beta_j X_j + \sum_{i<j}\sum \beta_{ij} X_i X_j + e \tag{5.33}$$

When interaction is significant, this model shows some curvature (twisting of the response surface plane). When curvature is not adequately modelled by Eq. (5.33), we consider a second order model

$$Y = \beta_0 + \sum_{j=1}^{K} \beta_j X_j + \sum_{i<j}\sum \beta_{ij} X_i X_j + \sum_{j=1}^{K} \beta_{jj} X_j^2 + e \tag{5.34}$$

where $\beta_{ij}$ represents pure second order or quadratic effect. Equation (5.34) is also called the second order response surface model.

In order to take curvature into account in $2^k$ designs, certain points are to be replicated which provides protection against curvature from second order model and also facilitate independent estimate of error. For this purpose some center points are added to the $2^k$ design. Figure 5.10 shows $2^2$ design with center points.

We know that in a $2^2$ design, the factorial points are (– , –), (–, +), (+, –) and (+, +). To these points if $n_C$ observations are added at the center point (0, 0), it reduces the number of experiments and at the same time facilitate to obtain experimental error. With this design usually one observation at each of the factorial point and $n_C$ observations at the center point are obtained. From the center points, error sum of squares is computed.

***Analysis of the design***

Let $\bar{Y}_C$ = average of the $n_C$ observations

$\bar{y}_F$ = average of the four factorial runs

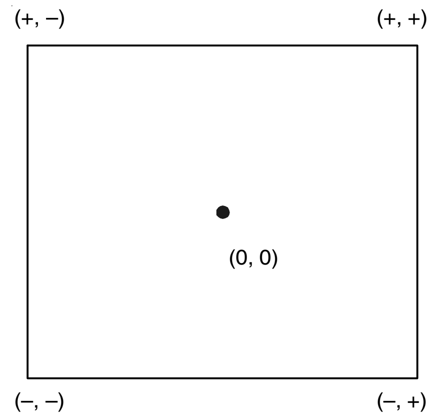



**FIGURE 5.10** The $2^2$ design with center points.

If $(\bar{y}_F - \bar{Y}_C)$ is small, then the center points lie on or near the plane passing through the factorial points and there will be no quadratic curvature.

If $(\bar{y}_F - \bar{Y}_C)$ is large, quadratic curvature will exist.

The sum of squares for factorial effects ($SS_A$, $SS_B$ and $SS_{AB}$) are computed as usual using contrasts. The sum of squares for quadratic effect is given by

$$SS_{\text{pure quadratic}} = \frac{n_F n_C (\bar{y}_F - \bar{y}_C)^2}{n_F + n_C} \text{ has one degree of freedom} \tag{5.35}$$

where $n_F$ is number of factorial design points.

This can be tested with error term. The error sum of squares is computed from the centre points only where we have replications.

$$SS_e = \sum_{i=1}^{nc} (Y_i - \bar{Y}_C)^2, \text{ where } y_i \text{ is } i\text{th data of center point} \tag{5.36}$$

$$= \sum y_i^2 - n_C \bar{y}_C^2 \tag{5.37}$$

**ILLUSTRATION 5.4**

**The $2^2$ Design with Center Points**

A chemist is studying the effect of two different chemical concentrations ($A$ and $B$) on the yield produced. Since the chemist is not sure about the linearity in the effects, he has decided to use a $2^2$ design with a single replicate augmented by three center points as shown in Figure 5.11.

The factors and their levels are as follows.

| *Factor* | *Low level* | *Center* | *High level* |
|---|---|---|---|
| *A* | 10 | 12.5 | 15 |
| *B* | 20 | 22.5 | 25 |

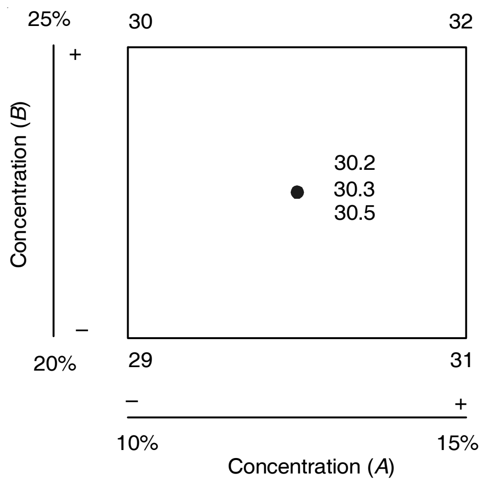



**FIGURE 5.11** The $2^2$ design with three center points for Illustration 5.4.

Suppose the data collected from the experiment at the four factorial points and the three center points are as given below:

| *Factorial points* | | | | *Center points* |
|---|---|---|---|---|
| (–1, –1) | (+1, –1) | (–1, +1) | (+1, +1) | 30.2, 30.3, 30.5 |
| 29 | 31 | 30 | 32 | |

***Data analysis:*** The sum of squares for the main and interaction effects is calculated using the contrasts.

$$SS_A = \frac{(C_A)^2}{n2^k} = \frac{[a + ab + (-1) - b]^2}{1 \times 2^2}$$

$$= \frac{(31 + 32 - 29 - 30)^2}{4} = 4.00$$

$$SS_B = \frac{(C_B)^2}{n2^k} = \frac{(b + ab - 1 - a)^2}{1 \times 2^2}$$

$$= \frac{(30 + 32 - 29 - 31)^2}{4} = 1.00$$

$$SS_{AB} = \frac{(C_{AB})^2}{n2^k} = \frac{[ab + (1) - a - b]^2}{4}$$

$$= \frac{(32 + 29 - 31 - 30)}{4} = 0.00$$

$$SS_{\text{pure quadratic}} = \frac{n_F n_C (\bar{y}_F - \bar{y}_C)^2}{n_F + n_C} \quad \text{[Eq. (5.35)]}$$

$$\bar{y}_F = \frac{(29 + 31 + 30 + 32)}{4} = 30.5$$

$$\bar{Y}_C = \frac{(30.2 + 30.3 + 30.5)}{3} = 30.3333$$

Substituting the respective values in Eq. (5.35), we obtain

$$SS_{\text{pure quadratic}} = 0.0495$$

The error sum of squares are computed from the center points and is given by

$$SS_e = \sum y_i^2 - n_C \bar{y}_C^2 \quad \text{[Eq. (5.37)]}$$

$$= 2760.38 - 2760.3333$$

$$= 0.0467$$

These computations are summarized in Table 5.14.

**TABLE 5.14** ANOVA for Illustration 5.4 ($2^2$ design with center points)

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean squares* | $F_0$ |
|---|---|---|---|---|
| *A* (Con. 1) | 4.0000 | 1 | 4.0000 | 170.9 |
| *B* (Con. 2) | 1.0000 | 1 | 1.0000 | 42.7 |
| *AB* | 0.0000 | 1 | 0.0000 | — |
| Pure quadratic | 0.0495 | 1 | 0.0495 | 2.1 |
| Error | 0.0467 | 2 | 0.0234 | |
| Total | 5.0962 | | | |

At 5% level of significance both factors *A* and *B* are significant. The interaction effect and the quadratic effects are insignificant. Hence, for this problem the first order regression model is sufficient.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + e \tag{5.38}$$

It is to be noted that the addition of center points do not affect the effect estimates.

If quadratic effect is significant, quadratic term is required in the model in which case a second order model has to be used.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + e \tag{5.39}$$

In this second order model Eq. (5.39), we have six unknown parameters. Whereas in the $2^2$ design with one center point, we have only five independent runs (experiments). In order to obtain solution to this model Eq. (5.39), it is required to use four axial runs in the $2^2$ design with one center point. This type of design is called a *central composite design*. The central composite design is discussed in Chapter 8.

## PROBLEMS

**5.1** A study was conducted to determine the effect of temperature and pressure on the yield of a chemical process. A $2^2$ design was used and the following data were obtained (Table 5.15).

**TABLE 5.15** Data for Problem 5.1

| *Pressure* | *Temperature* | |
|---|---|---|
| | 1 | 2 |
| 1 | 29, 33 | 46, 48 |
| 2 | 24, 22 | 48, 44 |

(a) Compute the factor and interaction effects.
(b) Test the effects using ANOVA and comment on the results.

**5.2** An Industrial engineer has conducted a study on the effect of speed ($A$), Tool type ($B$) and feed ($C$) on the surface roughness of machined part. Each factor was studied at two levels and three replications were obtained. The results obtained are given in Table 5.16.

**TABLE 5.16** Data for Problem 5.2

| *Treatment* | *Design* | | | *Roughness* | | |
|---|---|---|---|---|---|---|
| *combination* | $A$ | $B$ | $C$ | $R_1$ | $R_2$ | $R_3$ |
| (1) | – | – | – | 55 | 53 | 73 |
| *a* | + | – | – | 86 | 83 | 66 |
| *b* | – | + | – | 59 | 58 | 51 |
| *ab* | + | + | – | 75 | 65 | 65 |
| *c* | – | – | + | 61 | 61 | 53 |
| *ac* | + | – | + | 73 | 75 | 73 |
| *bc* | – | + | + | 50 | 55 | 49 |
| *abc* | + | + | + | 65 | 61 | 77 |

(a) Estimate the factor effects.
(b) Analyse the data using Analysis of Variance and comment on the results.

**5.3** Suppose the experimenter has conducted only the first replication ($R_1$) in Problem 5.2. Estimate the factor effects. Use half-normal plot to identify significant effects.

**5.4** In a chemical process experiment, the effect of temperature ($A$), pressure ($B$) and catalyst ($C$) on the reaction time has been studied and obtained data are given in Table 5.17.

**TABLE 5.17** Data for Problem 5.4

| *Treatment combination* | *Design* *A* | *B* | *C* | *Reaction time* $R_1$ | $R_2$ |
|---|---|---|---|---|---|
| (1) | – | – | – | 16 | 19 |
| *a* | + | – | – | 17 | 14 |
| *b* | – | + | – | 15 | 12 |
| *ab* | + | + | – | 16 | 13 |
| *c* | – | – | + | 21 | 27 |
| *ac* | + | – | + | 19 | 16 |
| *bc* | – | + | + | 20 | 24 |
| *abc* | + | + | + | 24 | 29 |

(a) Estimate the factor effects.
(b) Analyse the data using Analysis of Variance and comment on the results.
(c) Develop a regression model to predict the reaction time.
(d) Perform residual analysis and comment on the model adequacy.

**5.5** An experiment was conducted with four factors *A*, *B*, *C* and *D*, each at two levels. The following response data have been obtained (Table 5.18).

**TABLE 5.18** Data for Problem 5.5

| *Treatment combination* | *Response* $R_1$ | $R_2$ | *Treatment combination* | *Response* $R_1$ | $R_2$ |
|---|---|---|---|---|---|
| (1) | 41 | 43 | *d* | 47 | 45 |
| *a* | 24 | 27 | *ad* | 22 | 25 |
| *b* | 32 | 35 | *bd* | 35 | 33 |
| *ab* | 33 | 29 | *abd* | 37 | 35 |
| *c* | 26 | 28 | *cd* | 47 | 40 |
| *ac* | 31 | 29 | *acd* | 28 | 25 |
| *bc* | 36 | 32 | *bcd* | 37 | 35 |
| *abc* | 23 | 20 | *abcd* | 30 | 31 |

Analyse the data and draw conclusions.

**5.6** A $2^4$ factorial experiment was conducted with factors *A*, *B*, *C* and *D*, each at two levels. The data collected are given in Table 5.19.

**TABLE 5.19** Data for Problem 5.6

| *Treatment combination* | *Response* | *Treatment combination* | *Response* |
|---|---|---|---|
| (1) | 7 | *d* | 5 |
| *a* | 13 | *ad* | 20 |
| *b* | 8 | *bd* | 8 |
| *ab* | 11 | *abd* | 19 |
| *c* | 12 | *cd* | 14 |
| *ac* | 10 | *acd* | 16 |
| *bc* | 15 | *bcd* | 12 |
| *abc* | 10 | *abcd* | 18 |

(a) Obtain a normal probability plot of effects and identify significant effects.
(b) Analyse the data using ANOVA.
(c) Write down a regression model.

CHAPTER 6

# Blocking and Confounding in $2^k$ Factorial Designs

## 6.1 INTRODUCTION

We have already seen what is blocking principle. Blocking is used to improve the precision of the experiment. For example, if more than one batch of material is used as experimental unit in an experiment, we cannot really say whether the response obtained is due to the factors considered or due to the difference in the batches of material used. So, we treat batch of material as block and analyse the data. There may be another situation where one batch of material is not enough to conduct all the replications but sufficient to complete one replication. In this case, we treat each replication as a block. The runs in each block are randomized.

So, blocking is a technique for dealing with controllable nuisance variables. Blocking is used when resources are not sufficient to conduct more than one replication and replication is desired. There are two cases of blocking, i.e., replicated designs and unreplicated designs.

## 6.2 BLOCKING IN REPLICATED DESIGNS

If there are $n$ replicates, each replicate is treated as a block. Each replicate is run in one of the blocks (time periods, batches of raw materials etc.). Runs within the blocks are randomized.

**ILLUSTRATION 6.1**

Consider the chemical process experiment of $2^2$ design discussed in Chapter 5 (Illustration 5.1). The two replications of this experiment are given in Table 6.1.

**TABLE 6.1** Data from Illustration 6.1

| *Treatment* | *Replications* 1 | *Replications* 2 |
|---|---|---|
| (1) | 40 | 37 |
| *a* | 43 | 50 |
| *b* | 59 | 54 |
| *ab* | 37 | 43 |

Suppose one batch of material is just enough to run one replicate. Each replication is run in one block. Thus, we will have two blocks for this problem as shown in Figure 6.1. From Figure 6.1 the block totals are Block 1 ($B_1$) = 179 and Block 2 ($B_2$) = 184.

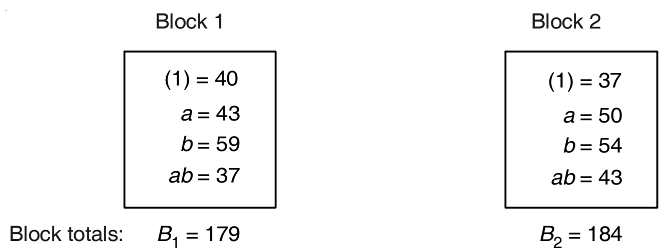



**FIGURE 6.1** Blocking a replicated design.

The analysis of this design is same as that of $2^2$ design. Here, in addition to effect sum of squares, we will have the block sum of squares.

Grand total ($T$) = 179 + 184 = 363

$$\text{Correction factor (CF)} = \frac{T^2}{N}$$

$$SS_{\text{Blocks}} = \frac{B_1^2 + B_2^2}{n_B} - \text{CF, where } n_B \text{ is the number of observations in } B. \tag{6.1}$$

$$= \frac{(179)^2 + (184)^2}{4} - \frac{(363)^2}{8}$$

$$= 16474.25 - 16471.125 = 3.125$$

The analysis of variance for Illustration 6.1 is given in Table 6.2. Note that the effect sum of squares are same as in Illustration 5.1.

**TABLE 6.2** Chemical process experiment with two blocks

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ | *Significance at 5%* |
|---|---|---|---|---|---|
| Blocks | 3.125 | 1 | 3.125 | | — |
| Temperature | 36.125 | 1 | 36.125 | 1.92 | Not significant |
| Pressure | 66.125 | 1 | 66.125 | 3.52 | Not significant |
| Interaction | 300.125 | 1 | 300.125 | 15.97 | Significant |
| Error | 56.375 | 3 | 18.79 | | |
| Total | 461.875 | 7 | | | |

It is to be noted that the conclusions from this experiment is same as that obtained in $2^2$ design earlier. This is because the block effect is very small. Also, note that we are not interested in testing the block effect.

## 6.3 CONFOUNDING

There are situations in which it may not be possible to conduct a complete replicate of a factorial design in one block where the block might be one day, one homogeneous batch of raw material etc. Confounding is a design technique for arranging a complete replication of a factorial experiment in blocks. So the block size will be smaller than the number of treatment combination in one replicate. That is, one replicate is split into different blocks. In this technique certain treatment effects are confounded (indistinguishable) with blocks. Usually, higher order interactions are confounded with blocks. When an effect is confounded with blocks, it is indistinguishable from the blocks. That is, we will not be able to say whether it is due to factor effect or block effect. Since, higher order interactions are assumed as negligible, they are usually confounded with blocks.

In $2^k$ factorial design, $2^m$ incomplete blocks are formed, where $m < k$. Incomplete block means that each block does not contain all treatment combinations of a replicate. These designs results in two blocks, four blocks, eight blocks and so on as given in Table 6.3.

**TABLE 6.3** Number of blocks possible in $2^k$ factorial design

| $k$ | $2^k$ | $m$ (*possible value*) | *No. of blocks* $2^m$ | *No. of treatments in each block* |
|---|---|---|---|---|
| 2 | 4 | 1 | 2 | 2 |
| 3 | 8 | 1 | 2 | 4 |
| | | 2 | 4 | 2 |
| 4 | 16 | 1 | 2 | 8 |
| | | 2 | 4 | 4 |
| | | 3 | 8 | 2 |

When the number of factors are large (screening experiments), there may be a constraint on the resources required to conduct even one complete replication. In such cases, confounding designs are useful.

## 6.4 THE $2^k$ FACTORIAL DESIGN IN TWO BLOCKS

Consider a $2^2$ factorial design with factors *A* and *B*. A single replicate of this design has 4 treatment combinations [(1), *a*, *b*, *ab*]. Suppose to conduct this experiment certain chemical is required and available in small batch quantities. If each batch of this chemical is just sufficient to run only two treatment combinations, two batches of chemical is required to complete one replication. Thus, the two batches of chemical will be the two blocks and we assign two of the four treatments to each block.

### 6.4.1 Assignment of Treatments to Blocks Using Plus–Minus Signs

The plus-minus table (Table 5.2) used to develop contrasts can be used to determine which of the treatments to be assigned to the two blocks. For our convenience Table 5.2 is reproduced as Table 6.4.

**TABLE 6.4** Plus–minus signs for $2^2$ design

| *Treatment combination* | *Factorial effect* | | | |
|---|---|---|---|---|
| | *I* | *A* | *B* | *AB* |
| (1) | + | – | – | + |
| *a* | + | + | – | – |
| *b* | + | – | + | – |
| *ab* | + | + | + | + |

Now we have to select the effect to be confounded with the blocks. As already pointed out, we confound the highest order interaction with the blocks. In this case, it is the *AB* interaction. In the plus–minus table above, under the effect *AB*, the treatments corresponding to the +ve sign are (1) and *ab*. These two treatments are assigned to one block and the treatments corresponding to the –ve sign (*a* and *b*) are assigned to the second block. The design is given in Figure 6.2.

| Block 1 | Block 2 |
|---|---|
| (1)<br>*ab* | *a*<br>*b* |

**FIGURE 6.2** A $2^2$ design in two blocks with *AB* confounded.

The treatments within the block are selected randomly for experimentation. Also which block to run first is randomly selected. The block effect is given by the difference in the two block totals. This is given below:

$$\begin{aligned}\text{Block effect} &= [(1) + ab] - [a + b]\\ &= (1) + ab - a - b \qquad (6.2)\end{aligned}$$

Note that the block effect $(1) + ab - a - b$ is equal to the contrast used to estimate the *AB* interaction effect. This indicates that the interaction effect is indistinguishable from the block effect. That is, *AB* is confounded with the blocks. From this it is evident that the plus–minus table can be used to create two blocks for any $2^k$ factorial designs. This concept can be extended to confound any effect with the blocks. But the usual practice is to confound the highest order interaction with the blocks. As another example consider a $2^3$ design to run in two blocks. Let the three factors be *A*, *B* and *C*. Suppose, we want to confound *ABC* interaction with the blocks. From the signs of plus–minus given in Table 6.5 (same as Table 5.3), we assign all the treatments corresponding to +ve sign under the effect *ABC* to block 1 and the others to block 2. The resulting design is shown in Figure 6.3.

The treatment combinations within each block are run in a random manner.

**TABLE 6.5** Plus–minus signs for $2^3$ design

| *Treatment combination* | *Factorial effect* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | *I* | *A* | *B* | *AB* | *C* | *AC* | *BC* | *ABC* |
| (1) | + | – | – | + | – | + | + | – |
| *a* | + | + | – | – | – | – | + | + |
| *b* | + | – | + | – | – | + | – | + |
| *ab* | + | + | + | + | – | – | – | – |
| *c* | + | – | – | + | + | – | – | + |
| *ac* | + | + | – | – | + | + | – | – |
| *bc* | + | – | + | – | + | – | + | – |
| *abc* | + | + | + | + | + | + | + | + |

| Block 1 | Block 2 |
|---|---|
| *a* | (1) |
| *b* | *ab* |
| *c* | *ac* |
| *abc* | *bc* |

**FIGURE 6.3** A $2^3$ design in two blocks with *ABC* confounded.

### 6.4.2 Assignment of Treatments Using Defining Contrast

This method makes use of a defining contrast

$$L = \lambda_1 X_1 + \lambda_2 X_2 + \lambda_3 X_3 + \lambda_4 X_4 + \cdots + \lambda_K X_K \tag{6.3}$$

where

$X_i$ = level of $i$th factor appearing in a particular treatment combination.
$\lambda_i$ = exponent on the $i$th factor in the effect to be confounded.

In $2^k$ system we will have $\lambda_i = 0$ or 1 and also $X_i = 0$ (low level) or $X_i = 1$ (high level). Suppose in a $2^3$ design, *ABC* is confounded with the blocks.

Let $X_1 \to A$, $X_2 \to B$ and $X_3 \to C$. Since all the three factors appear in the confounded effect, we will have $\lambda_1 = \lambda_2 = \lambda_3 = 1$. And the defining contrast corresponding to *ABC* is

$$L = X_1 + X_2 + X_3$$

On the other hand, if we want to confound *AC* with the blocks, we will have $\lambda_1 = \lambda_3 = 1$ and $\lambda_2 = 0$.

And the defining contrast corresponding to *AC* would be

$$L = X_1 + X_3$$

We assign the treatments that produce the same value of $L$ (mod 2) to the same block. Since the possible value of $L$ (mod 2) are 0 or 1, we assign all treatments with $L$ (mod 2) = 0 to one block and all treatments with $L$ (mod 2) = 1 to the second block. Thus, all the treatments will be assigned to two blocks. This procedure is explained as follows:

Suppose we want to design a confounding scheme for a $2^3$ factorial in two blocks with *ABC* confounded with the blocks. The defining contrast for this case is

$$L = X_1 + X_2 + X_3 \tag{6.4}$$

Using (0, 1) notation, we have to find $L$ (mod 2) for all the treatments.

$$\begin{aligned}
(1) &: L = 1(0) + 1(0) + 1(0) = 0 = 0 \text{ (mod 2)}\\
a &: L = 1(1) + 1(0) + 1(0) = 1 = 1 \text{ (mod 2)}\\
b &: L = 1(0) + 1(1) + 1(0) = 1 = 1 \text{ (mod 2)}\\
ab &: L = 1(1) + 1(1) + 1(0) = 2 = 0 \text{ (mod 2)}\\
c &: L = 1(0) + 1(0) + 1(1) = 1 = 1 \text{ (mod 2)}\\
ac &: L = 1(1) + 1(0) + 1(1) = 2 = 0 \text{ (mod 2)}\\
bc &: L = 1(0) + 1(1) + 1(1) = 2 = 0 \text{ (mod 2)}\\
abc &: L = 1(1) + 1(1) + 1(1) = 3 = 1 \text{ (mod 2)}
\end{aligned}$$

Thus (1), *ab*, *ac*, *bc* are assigned to block 1 and *a, b, c, abc* are assigned to block 2. This assignment is same as the design (Figure 6.3) obtained using plus–minus signs of Table 6.5.

### 6.4.3 Data Analysis from Confounding Designs

Analysing data through ANOVA require experimental error. So, replication of the experiment is necessary. If replication is not possible due to constraint on resources, the data are analysed as in single replication case of any other factorial design. That is, significant effects are identified using normal probability plot of effects and the error sum of squares is obtained by pooling the sum of squares of insignificant effects. When replication is possible in the confounding design either complete confounding or partial confounding is used.

## 6.5 COMPLETE CONFOUNDING

Confounding the same effect in each replication is called completely/fully confounded design. Thus in this design, we will not be able to obtain the information about the confounded effect. For example, consider a $2^3$ design to be run in two blocks with *ABC* confounded with the blocks and the experimenter wants to take three replications. The resulting design is shown in Figure 6.4 and the ANOVA is given in Table 6.6.

| Replicate 1 | | Replicate 2 | | Replicate 3 | |
|---|---|---|---|---|---|
| Block 1 | Block 2 | Block 1 | Block 2 | Block 1 | Block 2 |
| *a* | (1) | *a* | (1) | *a* | (1) |
| *b* | *ab* | *b* | *ab* | *b* | *ab* |
| *c* | *ac* | *c* | *ac* | *c* | *ac* |
| *abc* | *bc* | *abc* | *bc* | *abc* | *bc* |

**FIGURE 6.4** A $2^3$ design with *ABC* confounded in three replicates.

**TABLE 6.6** Analysis of variance for the $2^3$ design with *ABC* confounded in three replicates

| *Source of variation* | *Degrees of freedom* |
|---|---|
| Replicate | 2 |
| Blocks (*ABC*) within replicate | 3 |
| *A* | 1 |
| *B* | 1 |
| *C* | 1 |
| *AB* | 1 |
| *AC* | 1 |
| *BC* | 1 |
| Error | 12 |
| Total | 23 |

## 6.6 PARTIAL CONFOUNDING

In complete confounding, the information on the confounded effect cannot be retrieved because the same effect is confounded in all the replicates (Figure 6.4). Whenever replication of a confounded design is possible, instead of confounding the same effect in each replication, different effect is confounded in different replications. Such design is called *partial confounding*. For example, consider the design shown in Figure 6.5. There are three replicates. Interaction *ABC* is confounded in Replication 1, *AB* is confounded in Replication 2 and *AC* is confounded in Replication 3. In this design we can test all the effects. We can obtain information about *ABC* from the data in Replicates 2 and 3, information about *AB* from Replicates 1 and 3 and information about *AC* from Replicates 1 and 2. Thus in this design we say that 2/3 of the relative information about confounded effects can be retrieved. Table 6.7 gives the analysis of variance for the partially confounded $2^3$ design with three replicates shown in Figure 6.5.

**TABLE 6.7** ANOVA for the partially confounded $2^3$ design with three replicates

| *Source of variation* | *Degree of freedom* |
|---|---|
| Replicates | 2 |
| Blocks within replicate<br>(*ABC* in Replicate 1 + *AB* in Replicate 2 + *AC* in Replicate 3) | 3 |
| *A* | 1 |
| *B* | 1 |
| *C* | 1 |
| *AB* (from Replicates 1 and 3) | 1 |
| *AC* (from Replicates 1 and 2) | 1 |
| *BC* | 1 |
| *ABC* (from Replicates 2 and 3) | 1 |
| Error | 11 |
| Total | 23 |

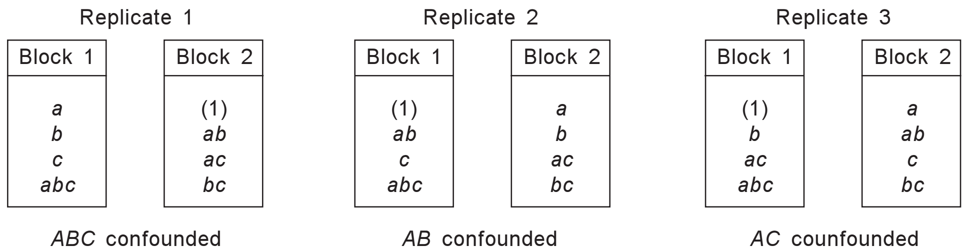



**FIGURE 6.5** Partial confounding of the $2^3$ design.

## ILLUSTRATION 6.2

**Partial Confounding**

A study was conducted to determine the effect of temperature ($A$), pressure ($B$) and stirring rate ($C$) on the yield of a chemical process. Each factor was studied at two levels. As each batch of raw material was just enough to test four treatment combinations, each replicate of the $2^3$ design was run in two blocks. The two replicates were run with $ABC$ confounded in Replicate 1 and $AC$ confounded in Replicate 2. The design and the coded data are shown in Figure 6.6.

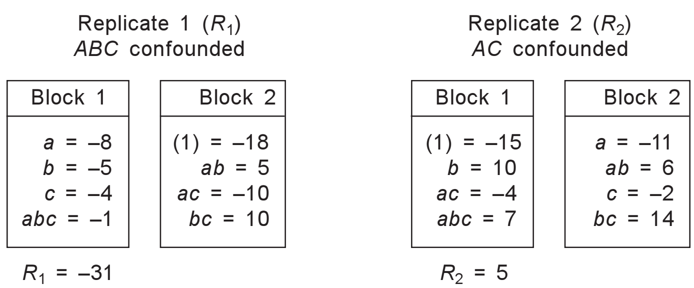



**FIGURE 6.6** Data for Illustration 6.2.

The sum of squares of $A$, $B$, $C$, $AB$, and $BC$ are computed using data from both the replications. Since $ABC$ is confounded in Replicate 1, $SS_{ABC}$ is computed from Replicate 2 only. Similarly, $SS_{AC}$ is computed from Replicate 1 only. And the sum of squares of replication ($SS_{\text{Rep}}$) is computed using the replication totals. The response totals for all the treatments are given in Table 6.8.

**TABLE 6.8** Response totals for Illustration 6.2

| Treatment | (1) | *a* | *b* | *ab* | *c* | *ac* | *bc* | *abc* |
|---|---|---|---|---|---|---|---|---|
| Total | –33 | –19 | 5 | 11 | –6 | –14 | 24 | 6 |

The sum of squares can be computed using the contrasts. These can be derived from Table 5.3.

$$SS_A = \frac{[-(1) + a - b + ab - c + ac - bc + abc]^2}{(n 2^k)}$$

$$= \frac{[-(-33) + (-19) - 5 + 11 - (-6) + (-14) - 24 + 6]^2}{[2 * 8]}$$

$$= \frac{(-6)^2}{16} = 2.25$$

$$SS_B = \frac{[-(1) - a + b + ab - c - ac + bc + abc]^2}{(n * 2^k)}$$

$$= \frac{[-(-33) - (-19) + 5 + 11 - (-6) - (-14) + 24 + 6]^2}{16}$$

$$= \frac{(118)^2}{16} = 870.25$$

$$SS_C = \frac{[-(1) - a - b - ab + c + ac + bc + abc]^2}{16}$$

$$= \frac{[-(-33) - (-19) - 5 - 11 + (-6) + (-14) + 24 + 6]^2}{16}$$

$$= \frac{(46)^2}{16} = 132.25$$

$$SS_{AB} = \frac{[(1) - a - b + ab + c - ac - bc + abc]^2}{(n * 2^k)}$$

$$= \frac{[-33 - (99) - 5 + 11 - 6 - (-14) - 24 + 6]^2}{16}$$

$$= \frac{(-18)^2}{16} = 20.25$$

$$SS_{BC} = \frac{[(1) + a - b - ab - c - ac + bc + abc]^2}{(n * 2^k)}$$

$$= \frac{[-33 - 19 - 5 - 11 - (-6) - (-14) + 24 + 6]^2}{16}$$

$$= \frac{(-18)^2}{16} = 20.25$$

$SS_{ABC}$ is computed using data from Replicate 2 only.

$$SS_{ABC} = \frac{[-(1) + a + b - ab + c - ac - bc + abc]^2}{(n * 2^k)}$$

$$= \frac{[-(-15) + (-11) + 10 - 6 + (-2) - (-4) - 14 + 7]^2}{(1 * 8)}$$

$$= \frac{(3)^2}{8} = 1.125$$

$SS_{AC}$ is computed using data from Replicate 1 only.

$$SS_{AC} = \frac{[(1) - a + b - ab - c + ac - bc + abc]^2}{n * 2^k}$$

$$= \frac{[-18 - (-8) - 5 - 5 - (-4) - 10 - 10 - 1]^2}{1 * 8}$$

$$= \frac{(-37)^2}{8} = 171.125$$

The replication sum of square is computed from the replication totals ($R_1$ = –31 and $R_2$ = 5) Grand total ($T$) = –26 and $N$ = 16.

$$\text{CF} = \frac{T^2}{N}$$

$$SS_{\text{Rep}} = \frac{R_1^2 + R_2^2}{n_{\text{Rep}}} - \text{CF} \tag{6.5}$$

where $n_{\text{Rep}}$ is the number of observation in the replication

$$SS_{\text{Rep}} = \frac{(-31)^2 + (5)^2}{8} - \frac{(-26)^2}{16}$$

$$= 123.25 - 42.25$$

$$= 81.00$$

The block sum of square is equal to the sum of $SS_{ABC}$ from Replicate 1 and $SS_{AC}$ from Replicate 2. It is found that

$$SS_{ABC} \text{ \{from Replicate 1\}} = 3.125$$

$$SS_{AC} \text{ \{from Replicate 2\}} = 10.125$$

Therefore,

$$SS_{\text{Block}} = 3.125 + 10.125 = 13.25$$

$$SS_{\text{Total}} = \sum_i \sum_j \sum_k \sum_l Y_{ijkl}^2 - \text{CF} \tag{6.6}$$

$$= [(-8)^2 + (-5)^2 + \cdots + (-2)^2 + (14)^2] - \frac{(-26)^2}{16}$$

$$= 1402 - 42.25 = 1359.75$$

$$SS_e = SS_{\text{Total}} - \Sigma\, SS \text{ of all the effects} + SS_{\text{Rep}} + SS_{\text{Block}}$$

$$= 1359.75 - 1311.75 = 48.00$$

The analysis of variance for Illustration 6.2 is given in Table 6.9.

**TABLE 6.9** ANOVA for $2^3$ design with partial confounding (Illustration 6.2)

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ | *Significance at* $\alpha = 5\%$ |
|---|---|---|---|---|---|
| Replicates | 1 | 81.00 | – | – | |
| Block within replications | 2 | 13.25 | – | – | |
| $A$ | 1 | 2.25 | 2.25 | 0.23 | Not significant |
| $B$ | 1 | 870.25 | 870.25 | 90.65 | Significant |
| $C$ | 1 | 132.25 | 132.25 | 13.78 | Significant |
| $AB$ | 1 | 20.25 | 20.25 | 2.11 | Not significant |
| $AC$ (from $R_1$) | 1 | 171.125 | 171.125 | 17.83 | Significant |
| $BC$ | 1 | 20.25 | 20.25 | 2.11 | Not significant |
| $ABC$ (from $R_2$) | 1 | 1.125 | 1.125 | .12 | Not significant |
| Error | 5 | 48.00 | 9.60 | | |
| Total | 15 | 1359.75 | | | |

From ANOVA it is found that the main effects $B$ and $C$ and the interaction $AC$ are significant and influence the process yield.

## 6.7 CONFOUNDING $2^k$ DESIGN IN FOUR BLOCKS

Confounding designs in four blocks are preferred when $k >= 4$. We obtain $2^{k-2}$ observations in each of these four blocks. We construct these designs using the method of defining contrast. We select two effects to be confounded with the blocks. Hence we will have two defining contrasts $L_1$ and $L_2$. Each contrast $L_1$ (mod 2) and $L_2$ (mod 2) will yield a specific pair of values for each treatment combination. In (0, 1) notation ($L_1$, $L_2$) will produce either (0, 0) or (0, 1) or (1, 0) or (1, 1). The treatment combinations producing the same values of ($L_1$, $L_2$) are assigned to the same block. Thus, all treatments with (0, 0) are assigned to first block, (0, 1) to the second block, (1, 0) to the third block and (1, 1) to the fourth block.

**ILLUSTRATION 6.3**

**Design of a $2^4$ Factorial Experiment in Four Blocks**

To obtain four blocks we select two effects to be confounded with the blocks. Suppose we want to confound *BC* and *AD* with the blocks. These effects will have two defining contrasts.

Let $A \rightarrow X_1$, $B \rightarrow X_2$, $C \rightarrow X_3$, $D \rightarrow X_4$

$$L_1 = X_2 + X_3 \tag{6.7}$$

$$L_2 = X_1 + X_4 \tag{6.8}$$

All the treatments are evaluated using $L_1$, $L_2$ as done in Section 6.4.2.

(1) : $L_1 = 1(0) + 1(0) = 0 \text{ (mod } 2 = 0)$

$L_2 = 1(0) + 1(0) = 0 \text{ (mod } 2 = 0)$

*a* : $L_1 = 0(\text{mod } 2 = 0)$ since $X_1$ does not appear in $L_1$

$L_2 = 1(1) + 0 = 1 \text{ (mod } 2 = 1)$

Similarly we can obtain $L_1$ (mod 2) and $L_2$ (mod 2) values for other treatments. These are given in Table 6.10 along with the assignment of blocks.

**TABLE 6.10** Assignment of treatments of $2^4$ design to four blocks

| *Treatment combination* | $L_1$ (mod 2) | $L_2$ (mod 2) | *Assignment* |
|---|---|---|---|
| (1) | 0 | 0 | Block 1 |
| *a* | 0 | 1 | Block 3 |
| *b* | 1 | 0 | Block 2 |
| *ab* | 1 | 1 | Block 4 |
| *c* | 1 | 0 | Block 2 |
| *ac* | 1 | 1 | Block 4 |
| *bc* | 0 | 0 | Block 1 |
| *abc* | 0 | 1 | Block 3 |
| *d* | 0 | 1 | Block 3 |
| *ad* | 0 | 0 | Block 1 |
| *bd* | 1 | 1 | Block 4 |
| *abd* | 1 | 0 | Block 2 |
| *cd* | 1 | 1 | Block 4 |
| *acd* | 1 | 0 | Block 2 |
| *bcd* | 0 | 1 | Block 3 |
| *abcd* | 0 | 0 | Block 1 |

The $2^4$ design in four blocks is shown in Figure 6.7.

| Block 1<br>$L_1 = 0$<br>$L_2 = 0$ | Block 2<br>$L_1 = 1$<br>$L_2 = 0$ | Block 3<br>$L_1 = 0$<br>$L_2 = 1$ | Block 4<br>$L_1 = 1$<br>$L_2 = 1$ |
|---|---|---|---|
| (1) | b | *a* | *ab* |
| *bc* | *c* | *d* | *ac* |
| *ad* | *abd* | *abc* | *bd* |
| *abcd* | *acd* | *bcd* | *cd* |

**FIGURE 6.7** The $2^4$ factorial design in four blocks with *BC* and *AD* confounded.

Note that there are 4 blocks with 3 degrees of freedom between them. But only two effects (*BC* and *AD* ) are confounded with the blocks and has one degree of freedom each. So, one more degree of freedom should be confounded. This should be the generalized interaction between the two effects *BC* and *AD*. This is equal to the product of *BC* and *AD* mod 2.

Thus, $BC * AD = ABCD$ will also be confounded with the blocks.

## 6.8 CONFOUNDING $2^k$ FACTORIAL DESIGN IN $2^m$ BLOCKS

Confounding $2^k$ factorial design method discussed in Section 6.7 can be extended to design $2^k$ factorial design in $2^m$ blocks ($m < k$), where each block contains $2^{k-m}$ treatments.

***Procedure:***

*Step* 1: Select $m$ independent effects to be confounded with the blocks.

*Step* 2: Construct $m$ defining contrasts $L_1, L_2, \ldots, L_m$ associated with the confounded effects.

This will create $2^m$ blocks, each containing $2^{k-m}$ treatments. We will have $2^m - m - 1$ generalized interactions from the $m$ independent effects selected for confounding which will also be confounded with the blocks.

Independent effect means that the effect selected for confounding should not become the generalized interaction of other effects. Also when selecting the $m$ independent effects for confounding, the resulting design should not confound the effects that are of interest to us.

The statistical analysis of these designs is simple. The sum of squares are computed as though no blocking is involved. Then the block sum of squares is found by adding the sum of squares of all the confounded effects.

The effects to be selected for generating the blocks is available in Montgomery (2003).

## PROBLEMS

**6.1** A study was conducted using a $2^3$ factorial design with factors *A*, *B* and *C*. The data obtained are given in Table 6.10.

**TABLE 6.10** Data for Problem 6.1

| *Treatment combination* | *Response* $R_1$ | $R_2$ |
|---|---|---|
| (1) | 15 | 12 |
| *a* | 17 | 23 |
| *b* | 34 | 29 |
| *ab* | 22 | 32 |
| *c* | 18 | 25 |
| *ac* | 5 | 6 |
| *bc* | 3 | 2 |
| *abc* | 12 | 18 |

Analyse the data assuming that each replicate ($R_1$ and $R_2$ ) as a block of one day.

**6.2** Consider the experiment described in Problem 5.2. Analyse this experiment assuming each replicate as a block.

**6.3** Consider the first replicate ($R_1$) of Problem 6.1. Suppose that this replicate could not be run in one day. Design an experiment to run these observations in two blocks with *ABC* confounded. Analyse the data.

**6.4** Consider the data in Problem 5.5. Using replicate 2, construct a design with two blocks of eight observations each confounding *ABCD*. Analyse the data.

**6.5** Consider the data in Problem 5.6. Construct a design with four blocks confounding *ABC* and *ABD* with the blocks. Analyse the data.

**6.6** Consider Problem 6.1. Suppose *ABC* is confounded in each replicate. Analyse the data and draw conclusions.

**6.7** Consider the data in Problem 6.1. Suppose *ABC* is confounded in replicate 1 and *BC* is confounded in replicate 2. Analyse the data.

**6.8** Consider the data from the second replication of Problem 5.5. Construct a design with four blocks confounding *ABC* and *ACD* with the blocks. Analyse the data.

# CHAPTER 7 Two-level Fractional Factorial Designs

## 7.1 INTRODUCTION

During process of product development one may have a large number of factors for investigation. As the number of factors increases, even in $2^k$ factorial designs, the number of experiments would be very high. For example, one complete replicate of $2^5$ designs would require 32 runs. Out of the 31 degrees of freedom only 5 degrees of freedom are associated with the main factors and 10 degrees of freedom account for two-factor interactions. The rest are associated with three-factor and other higher order interactions. If the experimenter assumes that certain higher order interactions are negligible, information on the main effects and lower order interactions can be obtained by conducting only a fraction of full factorial experiment. Even though these experiments have certain disadvantages, benefits are economies of time and other resources.

These fractional factorial designs are widely used in the industrial research to conduct screening experiments to identify factors which have large effects. Then detailed experiments are conducted on these identified factors.

When only a fraction of a complete replication is run, the design is called a *fractional replication* or *fractional factorial*.

## 7.2 THE ONE-HALF FRACTION OF THE $2^k$ DESIGN

For running a fractional factorial replication, a confounding scheme is used to run in two blocks and only one of the blocks is run. This forms the one-half fraction of $2^k$ design. As already pointed out, we confound the highest order interaction with the blocks. The confounded effect is called the *generator* of this fraction and sometimes it is referred to as a *word*.

Consider a $2^3$ design with three factors $A$, $B$ and $C$. One complete replication of this design consists of 8 runs. Suppose the experimenter cannot run all the 8 runs and can conduct only 4 runs. This becomes a one-half fraction of a $2^3$ design. Since this design consists of $2^{3-1} = 4$ treatment combinations, a one-half fraction of $2^3$ design, called $2^{3-1}$ design, where 3 denotes the number of factors and $2^{-1} = 1/2$ denotes the fraction. Table 7.1 is a plus–minus table for the $2^3$ design with the treatments rearranged.

**TABLE 7.1** Plus–minus table for $2^3$ design

| *Treatment combinations* | *Factorial effect* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | *I* | *A* | *B* | *AB* | *C* | *AC* | *BC* | *ABC* |
| *a* | + | + | – | – | – | – | + | + |
| *b* | + | – | + | – | – | + | – | + |
| *c* | + | – | – | + | + | – | – | + |
| *abc* | + | + | + | + | + | + | + | + |
| *ab* | + | + | + | + | – | – | – | – |
| *ac* | + | + | – | – | + | + | – | – |
| *bc* | + | – | + | – | + | – | + | – |
| (1) | + | – | – | + | – | + | + | – |

Suppose we select the treatment combinations corresponding to the + sign under $ABC$ as the one-half fraction, shown in the upper half of Table 7.1. The resulting design is the one-half fraction of the $2^3$ design and is called $2^{3-1}$ design (Table 7.2). Since the fraction is formed by selecting the treatments with +ve sign under $ABC$ effect, $ABC$ is called the *generator* of this fraction. Further, the identity element $I$ is always +ve, so we call

$I = +ABC$, the defining relation for the design

**TABLE 7.2** One-half fraction of $2^3$ design

| *Run* | *Factor* | | |
|---|---|---|---|
| | *A* | *B* | *C* |
| 1 | + | – | – |
| 2 | – | + | – |
| 3 | – | – | + |
| 4 | + | + | + |

The linear combination of observations used to estimate the effects of $A$, $B$ and $C$ are

$$l_A = 1/2(a - b - c + abc)$$
$$l_B = 1/2(-a + b - c + abc)$$
$$l_C = 1/2(-a - b + c + abc)$$

Similarly, the two factors interactions are estimated from

$$l_{AB} = 1/2(-a - b + c + abc)$$
$$l_{AC} = 1/2(-a + b - c + abc)$$
$$l_{BC} = 1/2(a - b - c + abc)$$

From the above, we find that

$$l_A = l_{BC}$$
$$l_B = l_{AC}$$
$$l_C = l_{AB}$$

Thus, we are getting confounding effects. $A$ is confounded with $BC$, $B$ is confounded with $AC$ and $C$ is confounded with $AB$. So, it is not possible to differentiate whether the effect is due to $A$ or $BC$ and so on. When two effects are confounded, we say that each is an alias of the other. That is, $A$ and $BC$, $B$ and $AC$, and $C$ and $AB$ aliases. This is denoted by

$$l_A \to A + BC$$
$$l_B \to B + AC$$
$$l_C \to C + AB$$

This is the alias structure for this design. Aliasing of effects is the price one must bear in these designs.

The alias structure can be determined by using the defining relation of the design, $I = ABC$. Multiplying the defining relation by any effect yields the alias of the effect. Suppose we want to find the alias for the effect $A$.

$$A \cdot I = A \cdot ABC = A^2BC$$

Since the square of any column ($A \times A$) is equal to the identity column $I$, we get

$$A = BC$$

Similarly, we can find the aliases for $B$ and $C$. This one-half fraction with $I = +ABC$ is called the *principle fraction.*

Suppose, we select the treatments corresponding to –ve sign under $ABC$ (Table 7.1) for forming the fraction. This will be a complementary one-half fraction. The defining relation for this design would be

$$I = -ABC$$

The alias structure for this fraction would be

$$l'_A \to A - BC$$
$$l'_B \to B - AC$$
$$l'_C \to C - AB$$

So, when we estimate $A$, $B$ and $C$ with this fraction we are, in fact, estimating $A - BC$, $B - AC$ and $C - AB$. In practice, any one fraction can be used. Both fractions put together form the $2^3$ design. If we run both fractions, we can obtain the de-aliased estimates of all effects by analyzing the eight runs as a $2^3$ design in two blocks. This can also be done by adding and subtracting the respective linear combinations of effects from the two fractions as follows:

$$1/2(l_A + l'_A) = 1/2(A + BC + A - BC) \to A$$
$$1/2(l_B + l'_B) = 1/2(B + AC + B - AC) \to B$$
$$1/2(l_C + l'_C) = 1/2(C + AB + C - AB) \to C$$
$$1/2(l_A - l'_A) = 1/2(A + BC - A + BC) \to BC$$
$$1/2(l_B - l'_B) = 1/2(B + AC - B + AC) \to AC$$
$$1/2(l_C - l'_C) = 1/2(C + AB - C + AB) \to AB$$

## 7.3 DESIGN RESOLUTION

The level of confounding of an experiment is called its *resolution*. So, design resolution is an indicator of the accuracy of the design in terms of the effect estimates. In full factorial design all the effects can be independently estimated. In fractional factorial design the effects are aliased with other effects. Suppose a main effect is aliased with a four-factor interaction. Its effect estimate can be said to be more accurate since a four-factor interaction can be assumed as negligible. On the other hand, if a main effect is aliased with a two-factor interaction, the estimate of main effect may not be accurate since a two-factor interaction may also exist. The design resolution deals with these aspects.

For the preceding $2^{3-1}$ design, the defining relation is $I = ABC \cdot ABC$ is called the *word* in the defining relation. The resolution of a two-level fractional factorial design is equal to the number of letters present in the smallest word in the defining relation. Thus, the resolution of $2^{3-1}$ design is Resolution III and is denoted as $2^{3-1}_{\text{III}}$ design. Usually, we represent the design resolution by a roman numeral. For example, if $I = ABCD$, it is of resolution IV design. Suppose $I = BDE = ACE = ABCD$, it is of resolution III design (appears a three-letter smallest word). We prefer to use highest resolution designs. The higher the resolution, the less restrictive shall be the assumption about the interactions aliased to be negligible.

The designs with resolution III, IV and V are considered as important. The characteristics of these designs are discussed as follows:

***Resolution III designs:*** A resolution III design does not have any main effect aliased with any other main effect, but are aliased with two-factor interactions. And two-factor interactions may be aliased with each other.

For example: A $2^{3-1}$ design with $I = ABC$

***Resolution IV designs:*** A resolution IV design does not have any main effect aliased with each other or two-factor interactions. Some two-factor interactions are aliased with other two-factor interactions.

For example: A $2^{4-1}$ design with $I = ABCD$

***Resolution V designs:*** In resolution V designs no main effect or two-factor interaction is aliased with any other main effect or two-factor interaction, but two-factor interactions are aliased with three-factor interactions.

For example: A $2^{5-1}$ design with $I = ABCDE$

## 7.4 CONSTRUCTION OF ONE-HALF FRACTION WITH HIGHEST RESOLUTION

The following steps discuss the construction of one-half fraction with highest resolution:

*Step* 1: Write down the treatment combinations for a full $2^{k-1}$ factorial with plus–minus signs. Each row of this is one run (experiment).

*Step* 2: Add the $k$th factor equal to the highest order interaction in the $2^{k-1}$ design and fill the plus–minus signs. And write down the resulting treatments for the fractional design.

This will be the $2^{k-1}$ fractional factorial design with the highest resolution.

Suppose we want to obtain a $2_{\text{III}}^{3-1}$ fractional factorial design,

*Step* 1: The treatments in $2^{3-1}$ full factorial ($2^2$ design) are (1), *a, b, ab* (Table 7.3).

**TABLE 7.3** $2^{3-1}$ full factorial

| | *A* | *B* |
|---|---|---|
| (1) | – | – |
| *a* | + | – |
| *b* | – | + |
| *ab* | + | + |

Replace the first column by runs.

*Step* 2: Add the *k*th factor *C* equal to the highest order interaction in $2^{k-1}$ design (*AB* interaction) (Table 7.4)

**TABLE 7.4** $2^{k-1}$ design

| *Run* | *A* | *B* | *C* = *AB* | *Treatment* |
|---|---|---|---|---|
| 1 | – | – | + | *c* |
| 2 | + | – | – | *a* |
| 3 | – | + | – | *b* |
| 4 | + | + | + | *abc* |

This is the required design which is a $2_{\text{III}}^{3-1}$ design with $I = ABC$.
Similarly we can obtain the other fraction with $I = -ABC$ (Table 7.5)

**TABLE 7.5** $2_{\text{III}}^{3-1}$, $I = -ABC$

| *Run* | *A* | *B* | *C* = – *AB* | *Treatment* |
|---|---|---|---|---|
| 1 | – | – | – | (1) |
| 2 | + | – | + | *ac* |
| 3 | – | + | + | *bc* |
| 4 | + | + | + | *abc* |

Note that these two fractions are the two blocks of a $2^3$ design, confounding *ABC* with the blocks.

## ILLUSTRATION 7.1

The surface finish of a machined component is being studied. Four factors, Speed (*A*), Feed (*B*), Depth of cut (*C*) and Type of coolant (*D*) are studied each at two levels. The coded data of the response (surface finish) obtained is given in Table 7.6.

**TABLE 7.6** Data for Illustration 7.1

| | $A_1$ | | | | $A_2$ | | | |
|---|---|---|---|---|---|---|---|---|
| | $B_1$ | | $B_2$ | | $B_1$ | | $B_2$ | |
| | $C_1$ | $C_2$ | $C_3$ | $C_4$ | $C_1$ | $C_2$ | $C_3$ | $C_4$ |
| $D_1$ | –18 | –17 | –9 | 1 | – 6 | 2 | 8 | 12 |
| $D_2$ | –11 | –12 | 5 | 1 | –2 | 9 | 8 | 10 |

The construction of a one-half fraction of the $2^{4-1}$ design with $I = ABCD$, its aliases, and the data analysis is discussed below.

The construction of one-half fraction of the $2^{4-1}$ design is given in Table 7.7. The response data (Table 7.6) of the treatments is also presented in Table 7.7.

**TABLE 7.7** The one-half fraction of the $2^{4-1}$ design with data

| *Run* | *A* | *B* | *C* | *D = ABC* | *Treatment* | *Response* |
|---|---|---|---|---|---|---|
| 1 | – | – | – | – | (1) | –18 |
| 2 | + | – | – | + | *ad* | –2 |
| 3 | – | + | – | + | *bd* | 5 |
| 4 | + | + | – | – | *ab* | 8 |
| 5 | – | – | + | + | *cd* | –12 |
| 6 | + | – | + | – | *ac* | 2 |
| 7 | – | + | + | – | *bc* | 1 |
| 8 | + | + | + | + | *abcd* | 10 |

The defining relation for the design is $I = ABCD$. The aliases are as follows:

$$l_A \to A + BCD$$
$$l_B \to B + ACD$$
$$l_C \to C + ABD$$
$$l_D \to D + ABC$$
$$l_{AB} \to AB + CD$$
$$l_{AC} \to AC + BD$$
$$l_{AD} \to AD + BC$$

The linear combination of observations (contrasts) used to estimate the effects and computation of sum of squares are as follows:

$$\begin{aligned} C_A &= -(1) + ad - bd + ab - cd + ac - bc + abcd \\ &= -(-18) - 2 - 5 + 8 - (-12) + 2 - 1 + 10 = 42 \\ C_B &= -(1) - ad + bd + ab - cd - ac + bc + abcd \\ &= -(-18) - (-2) + 5 + 8 - (-12) - 2 + 1 + 10 = 54 \end{aligned}$$

$$C_C = -(1) - ad - bd - ab + cd + ac + bc + abcd$$
$$= -(-18) - (-2) - 5 - 8 - 12 + 2 + 1 + 10 = 8$$
$$C_D = -(1) + ad + bd - ab + cd - ac - bc + abcd$$
$$= -(-18) - 2 + 5 - 8 - 12 + - 2 - 1 + 10 = 8$$
$$C_{AB} = +(1) - ad - bd + ab + cd - ac - bc + abcd$$
$$= -18 - (-2) - 5 + 8 - 12 - 2 - 1 + 10 = -18$$
$$C_{AC} = +(1) - ad + bd - ab - cd + ac - bc + abcd$$
$$= -18 - (-2)+ 5 - 8 - (-12) + 2 - 1 + 10 = 4$$
$$C_{AD} = +(1) + ad - bd - ab - cd - ac + bc + abcd$$
$$= -18 - 2 - 5 - 8 - (-12) - 2 + 1 + 10 = -12$$

Effect of factor/interaction $= \dfrac{C}{n2^{k-1}}$, where $C$ is the contrast

Sum of squares of a factor/interaction $= \dfrac{C^2}{n2^k}$ where $k = 4$ and $n = 1/2$ (one half of the replication)

The estimate of effects and sum of squares of all factors and interactions are summarized in Table 7.8.

**TABLE 7.8** Effects and sum of squares and alias structure for Illustration 7.1

| *Estimate of effect* | *Sum of squares* | *Alias structure* |
|---|---|---|
| $l_A = 10.5$ | $SS_A = 220.50$ | $l_A \rightarrow A + BCD$ |
| $l_B = 13.5$ | $SS_B = 364.50$ | $l_B \rightarrow B + ACD$ |
| $l_C = 2.0$ | $SS_C = 8.00$ | $l_C \rightarrow C + ABD$ |
| $l_D = 2.0$ | $SS_D = 8.00$ | $l_D \rightarrow D + ABC$ |
| $l_{AB} = -4.5$ | $SS_{AB} = 40.50$ | $l_{AB} \rightarrow AB + CD$ |
| $l_{AC} = 1.0$ | $SS_{AC} = 2.00$ | $l_{AC} \rightarrow AC + BD$ |
| $l_{AD} = -3.0$ | $SS_{AD} = 18.00$ | $l_{AD} \rightarrow AD + BC$ |

The normal plot and the half normal plot of the effects for Illustration 7.1 are shown in Figures 7.1 and 7.2 respectively.

From these two plots of effects (Figures 7.1 and 7.2), it is evident that the main effects $A$ and $B$ alone are significant. Hence, the sum of squares of the main effect $C$ and $D$ and the two-factor interactions are pooled into the error term. The analysis of variance is given in Table 7.9.

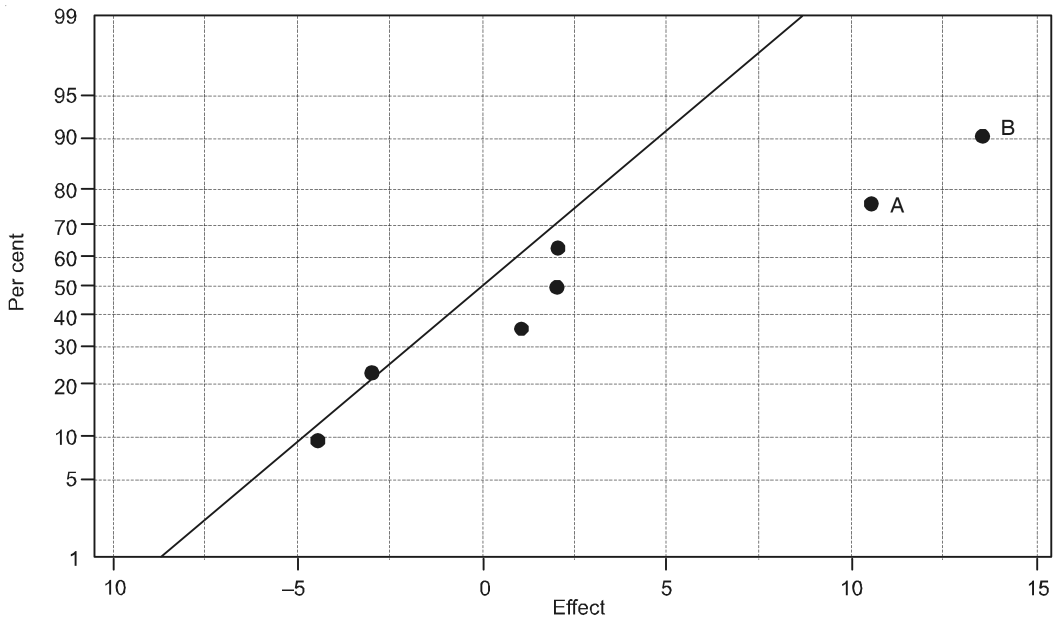



**FIGURE 7.1** Normal probability plot of the effects for Illustration 7.1.

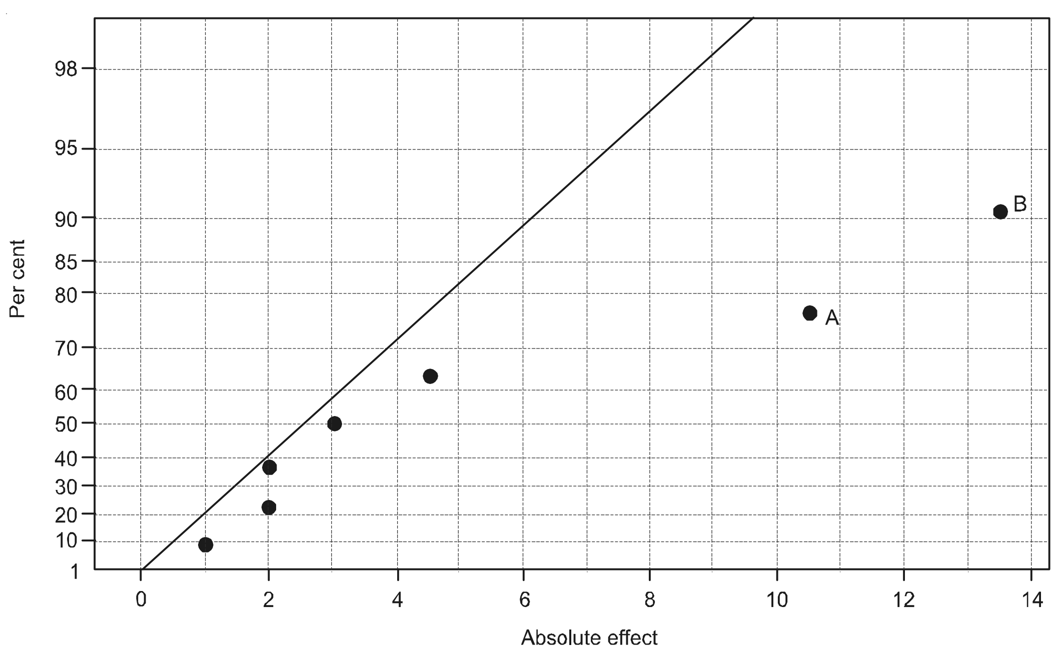



**FIGURE 7.2** Half-normal plot of the effects for Illustration 7.1.

**TABLE 7.9** Analysis of variance for Illustration 7.1

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ |
|---|---|---|---|---|
| *A* | 220.50 | 1 | 220.50 | 14.41 |
| *B* | 364.50 | 1 | 364.50 | 23.82 |
| Pooled error | 76.50 | 5 | 15.3 | |
| Total | 661.50 | 7 | | |

$F_{0.05,1,5} = 6.61$

From the analysis of variance, it is found that only the main factors speed (*A*) and feed (*B*) influence the surface finish.

## 7.5 THE ONE-QUARTER FRACTION OF $2^k$ DESIGN

When the number of factors to be investigated are large, one-quarter fraction factorial designs are used. This design will contain $2^{k-2}$ runs and is called $2^{k-2}$ *fractional design*. The design will have two generators and their generalized interaction in the defining relation.

***Construction of $2^{k-2}$ design:*** The following steps discuss the constructions of $2^{k-2}$ design:

*Step* 1: Write down the runs of a full factorial design with $k$–2 factors.

*Step* 2: Have two additional columns for other factors with appropriately chosen interactions involving the first $k$–2 factors and fill the plus–minus signs. And write down the resulting treatments which is the required $2^{k-2}$ fractional factorial design.

Suppose $k = 5$(*A*, *B*, *C*, *D* and *E*). Then, $k - 2 = 3$. The $2^{5-2}$ fractional factorial design is shown in Table 7.10.

**TABLE 7.10** The $2^{5-2}$ fractional factorial design

| *Run* | *A* | *B* | *C* | *D = AB* | *E = BC* | *Treatment* |
|---|---|---|---|---|---|---|
| 1 | – | – | – | + | + | *de* |
| 2 | + | – | – | – | + | *ae* |
| 3 | – | + | – | – | – | *b* |
| 4 | + | + | – | + | – | *abd* |
| 5 | – | – | + | + | – | *cd* |
| 6 | + | – | + | – | – | *ac* |
| 7 | – | + | + | – | + | *bce* |
| 8 | + | + | + | + | + | *abcde* |

The generators for this design ($2^{5-2}$) are $I = ABD$ and $I = BCE$. The generalized interaction of *ABD* and *BCE* is *ACDE*. So the complete defining relation for the design is $I = ABD = BCE = ACDE$. And the design resolution is III.

Note that, other fractions for this design can be obtained by changing the design generators like

$$D = -AB,\ E = -BC,\ E = \pm AC \text{ or } E = \pm ABC$$

The alias structure for the design given in Table 7.10 is given in Table 7.11.

**TABLE 7.11** The alias structure for the $2^{5-2}$ design ($I = ABD = BCE = ACDE$)

| |
|---|
| $A = BD = ABCE = CDE$ |
| $B = AD = CE = ABCDE$ |
| $C = ABCD = BE = ADE$ |
| $D = AB = BCDE = ACE$ |
| $E = ABDE = BC = ACD$ |
| $AC = BCD = ABE = DE$ |
| $AE = BDE = ABC = CD$ |

The analysis of data from these designs is similar to that of $2^{k-1}$ design discussed in Section 7.4.

## 7.6 THE $2^{k-m}$ FRACTIONAL FACTORIAL DESIGN

A $2^{k-m}$ fractional factorial containing $2^{k-m}$ runs is a $1/2^m$ fraction of the $2^k$ design or it is a $2^{k-m}$ fractional factorial design. The defining relation for this design contains $m$ generators selected initially and their $2^m-m-1$ generalized interactions. The alias structure is obtained by multiplying each effect by the defining relation. Each effect has $2^m-1$ aliases. For large value of $k$, we assume that higher order interactions (3 or more factors) are negligible. This simplifies the alias structure. The generators are selected such that the interested effects are not aliased with each other. Generally we select the generator such that the resulting $2^{k-m}$ design has the highest possible resolution. Guidelines are available for selecting the generator for the $2^{k-m}$ fractional designs (Montgomery 2003).

## 7.7 FRACTIONAL DESIGNS WITH SPECIFIED NUMBER OF RUNS

We can construct resolution III designs for studying up to $k = n - 1$ factors in $n$ runs, where $n$ is a multiple of 4. Thus we can have design with 4 runs (up to 3 factors), 8 runs (up to 7 factors), 16 runs (up to 15 factors), etc. These designs are called *saturated designs* ($k = n - 1$) since all columns will be assigned with factors. A design up to 3 factors with 4 runs is a $2_{\text{III}}^{3-1}$ design, which is one-half fraction of $2^3$ design. Studying 7 factors in 8 runs is a $2_{\text{III}}^{7-4}$ design (1/16 fraction of $2^7$ design). These designs can be constructed by the methods discussed in the preceding sections of this chapter. These designs are frequently used in the industrial experimentation.

## 7.8 FOLD-OVER DESIGNS

In resolution III designs, the main effects are aliased with two-factor interactions. In some experiments, the aliased main effect may not be significant and the interaction may really affect the response. This we do not know, since it is aliased with a main effect. In order to know this, further detailed experimentation is needed. One way is to design a separate experiment (full factorial or higher resolution design) and analyse. Alternatively, to utilize the data collected in the experiment already conducted along with a second experimental data, (to be conducted), a fold-over design is recommended. Suppose we have conducted an experiment using $2_{\text{III}}^{3-1}$ design with $+ABC$ as the generator. To this design, if we add the other fraction with $-ABC$ as the generator, we call such a design as *fold-over design*. If we add to a resolution III fractional design a second fraction in which the signs of all factors are reserved, resulting design is called a *full fold-over* or a *reflection design*. When we fold over a resolution III design, if the signs on the generators that have an odd number of letters is changed, we obtain a full fold-over design. Even if we change the signs of any one generator and add that fraction to another fraction, the resulting design is also a fold-over design. To a principal fraction if a second fraction is added with the signs of a main factor reversed, we can isolate this main effect and its two-factor interaction. The objective is to isolate the confounded effects in which we are interested. The example for a fold-over design is given in Table 7.12.

**TABLE 7.12** Full fold-over design for $2_{\text{III}}^{7-4}$ fractional design

**Original fraction ($F_1$)**

| *Run* | *A* | *B* | *C* | *D* = *AB* | *E* = *AC* | *F* = *BC* | *G* = *ABC* | *Treatment* |
|---|---|---|---|---|---|---|---|---|
| 1 | – | – | – | + | + | + | – | *def* |
| 2 | + | – | – | – | – | + | + | *afg* |
| 3 | – | + | – | – | + | – | + | *beg* |
| 4 | + | + | – | + | – | – | – | *abd* |
| 5 | – | – | + | + | – | – | + | *cdg* |
| 6 | + | – | + | – | + | – | – | *ace* |
| 7 | – | + | + | – | – | + | – | *bcf* |
| 8 | + | + | + | + | + | + | + | *abcdefg* |

**Fold-over fraction ($F_2$)**

| *Run* | *–A* | *–B* | *–C* | *–D* = *AB* | *–E* = *AC* | *–F* = *BC* | *–G* = *–ABC* | *Treatment* |
|---|---|---|---|---|---|---|---|---|
| 1′ | + | + | + | – | – | – | + | *abcg* |
| 2′ | – | + | + | + | + | – | – | *bcde* |
| 3′ | + | – | + | + | – | + | – | *acdf* |
| 4′ | – | – | + | – | + | + | + | *cefg* |
| 5′ | + | + | – | – | + | + | – | *abef* |
| 6′ | – | + | – | + | – | + | + | *bdfg* |
| 7′ | + | – | – | + | + | – | + | adeg |
| 8′ | – | – | – | – | – | – | – | (1) |

Note that we have to run the two fractions separately. The effects from the first fraction $F_1(l)$ and the second fraction $F_2(l')$ are estimated separately and are combined as $1/2(l + l')$ and $1/2(l - l')$ to obtain de-aliased effect estimates. The two fractions can be treated as two blocks with 8 runs each. The block effect can be obtained as the difference between the average response of the two fractions. The alias structure for the original fraction and its fold over can be simplified by assuming that the three-factor and higher order interactions as negligible.

The generators for the original fraction ($F_1$) and the fold-over fraction ($F_2$) are as follows:

$$\text{For } F_1: \quad I = ABD,\ I = ACE,\ I = BCF,\ I = ABCG \tag{7.1}$$

$$\text{For } F_2: \quad I = -ABD,\ I = -ACE,\ I = -BCF,\ I = ABCG \tag{7.2}$$

Note that the sign is changed only for the odd number of letter words of the generators in $F_1$ to obtain the fold-over design.

The complete defining relation for the original fraction ($F_1$) can be obtained by multiplying the four generators $ABD$, $ACE$, $BCF$ and $ABCG$ together two at a time, three at a time and all the four. Thus, for the original fraction, the defining relation is

$$I = \boldsymbol{ABD} = \boldsymbol{ACE} = \boldsymbol{BCF} = ABCG = BCDE = ACDF = \boldsymbol{CDG} = ABEF = \boldsymbol{BEG} = \boldsymbol{AFG} = \boldsymbol{DEF} = ADEG = CEFG = BDFG = \boldsymbol{ABCDEFG} \tag{7.3}$$

This consists eight words of odd length (bold faced) and seven words of even length. The defining relation for the fold-over fraction ($F_2$) can be obtained by changing the sign of words with odd number of letters in the defining relation of the original fraction ($F_1$). The remaining seven words with even number of letters will be same as in the defining relation of $F_1$. So the complete defining relation for $F_2$ is

$$I = -ABD = -ACE = -BCF = ABCG = BCDE = ACDF = -CDG = ABEF = -BEG = -AFG = -DEF = ADEG = CEFG = BDFG = -ABCDEFG \tag{7.4}$$

The fold-over design denoted by $F$ consists of $F_1$ and $F_2$. Since all the defining words of odd length have both +ve and –ve sign in $F$ do not impose any constraint on the factors and hence do not appear in the defining relation of $F$. The words of even length which are same in $F_1$ and $F_2$ are the defining words of the defining relation of $F$ which is

$$ABCG = BCDE = ACDF = ABEF = ADEG = CEFG = BDFG. \tag{7.5}$$

The design resolution of $F$ is IV. That is, the fold-over design ($F$) is a $2^{7-4}_{\text{IV}}$ design. Since it is a resolution IV design, all main effects are not aliased.

The aliases associated with the main effects and two factor interactions are given in Table 7.13. These are obtained from Eq. (7.3), neglecting higher order interactions.

**TABLE 7.13** The alias structure for the original fraction ($F_1$)

| |
|---|
| $l_A \rightarrow A + BD + CE + FG$ |
| $l_B \rightarrow B + AD + CF + EG$ |
| $l_C \rightarrow C + AE + BF + DG$ |
| $l_D \rightarrow D + AB + CG + EF$ |
| $l_E \rightarrow E + AC + BG + DF$ |
| $l_F \rightarrow F + BC + AG + DE$ |
| $l_G \rightarrow G + CD + BE + AF$ |

Similarly, the alias structure for the fold-over fraction is derived and is given in Table 7.14.

**TABLE 7.14** The alias structure for the fold-over fraction ($F_2$)

$$l'_A \rightarrow A - BD - CE - FG$$
$$l'_B \rightarrow B - AD - CF - EG$$
$$l'_C \rightarrow C - AE - BF - DG$$
$$l'_D \rightarrow D - AB - CG - EF$$
$$l'_E \rightarrow E - AC - BG - DF$$
$$l'_F \rightarrow F - BC - AG - DE$$
$$l'_G \rightarrow G - CD - BE - AF$$

By combining the fold-over fraction with the original as $1/2(l_i + l'_i)$, we get the estimates for all main factors and $1/2(l_i - l'_i)$ will lead to the estimates for the combined two-factor interactions.

## ILLUSTRATION 7.2

**Fold-over Design for $2^{7-4}_{\text{III}}$**

The problem is concerned with an arc welding process. The experiment was planned to find out the process parameter levels that maximize the welding strength. The following factors and levels have been used.

| *Factor* | *Low level* | *High level* |
|---|---|---|
| Type of welding rod (*A*) | X 100 | Y 200 |
| Weld material (*B*) | SS41 | SB35 |
| Thickness of welding material (*C*) | 8 mm | 12 mm |
| Angle of welded part (*D*) | 60°C | 70°C |
| Current (*E*) | 130 A | 150 A |
| Welding method (*F*) | Single | Weaving |
| Preheating (*G*) | No heating | 150°C |

These seven factors have been assigned to the seven columns of the fold-over design given in Table 7.12. The experiments in the two fractions have been run separately and the response (weld strength) for the first fraction and fold-over design (second fraction) measured is tabulated in the form of coded data in Table 7.15. The effects are estimated using contracts.

$$C_A = -def + afg - beg + abd - cdg + ace - bcf + abcdefg$$
$$= -47 - 9 - (-27) - 13 - (-16) - 22 - (-5) + 39 = -4$$

$$\text{Effect of } A = \frac{C_A}{n2^{k-1}} = \frac{-4}{1/16(2^6)} = -1.00$$

Similarly, we can estimate the effects of both the fractions. These are given in Table 7.16.

**TABLE 7.15** Weld strength data (coded value) for Illustration 7.2

| *First fraction* | | | *Second fraction (fold over)* | | |
|---|---|---|---|---|---|
| *Run* | *Treatment combination* | *Response* | *Run* | *Treatment combination* | *Response* |
| 1 | *def* | 47 | 1′ | *adcg* | –10 |
| 2 | *afg* | –9 | 2′ | *bcde* | 37 |
| 3 | *beg* | –27 | 3′ | *acdf* | –13 |
| 4 | *abd* | –13 | 4′ | *cefg* | –28 |
| 5 | *cdg* | –16 | 5′ | *abef* | –28 |
| 6 | *ace* | –22 | 6′ | *bdfg* | –13 |
| 7 | *bcf* | –5 | 7′ | *adeg* | 45 |
| 8 | *abcdefg* | 39 | 8′ | (1) | –7 |

**TABLE 7.16** Estimate of effects for Illustration 7.2

| *First fraction* | *Second fraction (fold over)* |
|---|---|
| $l'_A = 1.0 \rightarrow A + BD + CE + FG$ | $l'_A = 1.25 \rightarrow A - BD - CE - FG$ |
| $l'_B = -1.5 \rightarrow B + AD + CF + EG$ | $l'_B = -2.75 \rightarrow B - AD - CF - EG$ |
| $l'_C = -0.5 \rightarrow C + AE + BF + DG$ | $l'_C = -2.75 \rightarrow C - AE - BF - DG$ |
| $l'_D = 27.5 \rightarrow D + AB + CG + EF$ | $l'_D = 32.25 \rightarrow D - AB - CG - EF$ |
| $l'_E = 20.0 \rightarrow E + AC + BG + DF$ | $l'_E = 17.25 \rightarrow E - AC - BG - DF$ |
| $l'_F = 37.5 \rightarrow F + BC + AG + DE$ | $l'_F = -36.25 \rightarrow F - BC - AG - DE$ |
| $l'_G = 5.0 \rightarrow G + CD + BE + AF$ | $l'_G = 2.25 \rightarrow G - CD - BE - AF$ |

By combining the two fractions we can obtain the de-aliased effects for the main factors and the two-factor interactions together as given in Table 7.17.

**TABLE 7.17** De-aliased effects

| $i$ | $1/2(l_i + l'_i)$ | $1/2(l_i - l'_i)$ |
|---|---|---|
| $A$ | $A = 0.125$ | $BD + CE + FG = -1.125$ |
| $B$ | $B = -2.125$ | $AD + CF + EG = 0.625$ |
| $C$ | $C = -1.625$ | $AE + BF + DG = 1.125$ |
| $D$ | $D = 29.875$ | $AB + CG + EF = -2.375$ |
| $E$ | $E = 18.625$ | $AC + BG + DF = 1.375$ |
| $F$ | $F = 0.625$ | $BC + AG + DE = 36.875$ |
| $G$ | $G = 3.625$ | $CD + BE + AF = 1.375$ |

From Table 7.17, it is observed that factors $D$ and $E$ have large effects compared to others. The aliases for these two are

$$D = AB = CG = EF$$

$$E = AC = BG = DF$$

If it is assumed that the factors $A$, $B$, $C$ and $G$ do not have significant effect and not involve in any two-factor significant interactions. The above reduces to

$$D = EF$$

$$E = DF$$

The two-factor interaction combinations involving $EF$ and $DF$ have a small effect indicating that they are not significant (Table 7.17). But, the interaction $BC + AG + DE$ has a large effect. Since $BC$ and $AG$ can be assumed as negligible, we can conclude that the interaction $DE$ is significant. Hence, the optimal levels for $D$ and $E$ should be determined considering the response of $DE$ interaction combinations ($D_1E_1$, $D_1E_2$, $D_2E_1$ and $D_2E_2$). Since the other factors are insignificant, any level (low or high) can be used. But usually, we select the level for insignificant factors based on cost and convenience for operation.

Average response for $DE$ interaction combinations is

$$D_1E_1\text{: } 1/2(-9 + -5) = -7.0$$
$$D_1E_2\text{: } 1/2(-27 + -22) = -24.5$$
$$D_2E_1\text{: } 1/2(-13 + -16) = -14.5$$
$$D_2E_2\text{: } 1/2(47 + 39) = 43.0$$

The objective is to maximize the weld strength. So, the combination $D_2E_2$ is selected. For other factors, the levels are fixed as discussed above. Thus, the fold-over designs is useful to arrive at the optimal levels.

## PROBLEMS

**7.1** Consider the first replicate ($R_1$) of Problem 5.5. Suppose it was possible to run a one-half fraction of the $2^4$ design. Construct the design and analyse the data.

**7.2** Consider Problem 5.6. Suppose it was possible to run a one-half fraction of the $2^4$ design. Construct the design and analyse the data.

**7.3** Construct a $2^{5-2}$ design with $ACE$ and $BDE$ as generators. Determine the alias structure.

**7.4** Construct a $2^{7-2}$ design. Use $F = ABCD$ and $G = ABDE$ as generators. Write down the alias structure. Outline the analysis of variance table. What is the resolution of this design?

**7.5** Design a $2^{6-2}$ fractional factorial with $I = ABCE$ and $I = BCDF$ as generators. Give the alias structure.

# CHAPTER 8 Response Surface Methods

## 8.1 INTRODUCTION

In experimental investigation we study the relationship between the input factors and the response (output) of any process or system. The purpose may be to optimize the response or to understand the system. If the input factors are quantitative and are a few, Response Surface Methodology (RSM) can be used to study the relationship. Suppose we want to determine the levels of temperature ($X_1$) and pressure ($X_2$) that maximize the yield ($Y$) of a chemical process. That is, the process yield is a function of the levels of temperature and pressure which can be represented as

$$Y = f(X_1, X_2) + e$$

where $e$ is the observed error in the response $Y$.

If the expected response $E(Y) = f(X_1, X_2) = \eta$, then the surface is represented by

$$\eta = f(X_1, X_2)$$

An example of a response surface is shown in Figure 8.1.

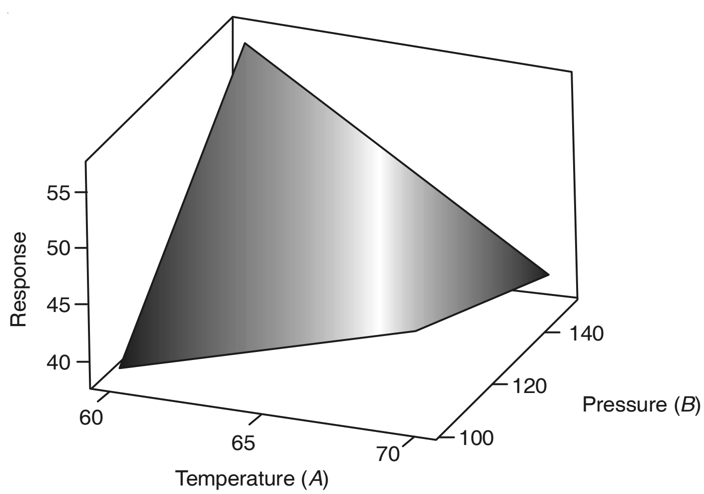



**FIGURE 8.1** A response surface.

In RSM, a sequential experimentation strategy is followed. This involves the search of input factor space by using first-order experiment followed by a second-order experiment. The first order equation is given by

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + e \tag{8.1}$$

If there is curvature in the system, higher order polynomial, usually a second-order model given below is used.

$$Y = \beta_0 + \sum_{i=1}^{K} \beta_i X_i + \sum_{i=1}^{K} \beta_{ii} X_i^2 + \sum_{i<j}\sum \beta_{ij} X_i X_j + e \tag{8.2}$$

The method of least squares is used to estimate the parameters (discussed in Chapter 2). The response surface analysis is then performed using the fitted surface. The objective of RSM is to determine the optimum operating conditions. This analysis is generally performed using computer software.

The model parameters can be estimated effectively if proper experimental designs are used to collect data. Such designs are called *response surface designs*. Some of the efficient designs used in second-order experiment are the Central Composite Design (CCD) and the Box–Behnken Design.

## 8.2 RESPONSE SURFACE DESIGNS

Selecting an appropriate design for fitting and analysing the response surface is important. These designs should allow minimum number of experimental runs but provide complete information required for fitting the model and checking model adequacy.

### 8.2.1 Designs for Fitting First-order Model

Suppose we want to fit a first-order model in $k$ variables

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + e \tag{8.3}$$

These designs include $2^k$ factorial and fractions of $2^k$ series in which main effects are not aliased with each other. Here the low and high levels of the factors are denoted by ±1 notation. We will not obtain error estimate in these designs unless some runs are replicated. A common method of including replications is to take some observations at the center point. This design has already been discussed in Chapter 5.

### 8.2.2 Central Composite Design (CCD)

The central composite design for $k$ factors include:

(i) $n_F$ factorial points (corner/cube points) each at two levels indicated by (–1, +1)
(ii) $n_C$ centre points indicated by 0.
(iii) $2k$ axial points (star points)

In selecting a CCD, the following three issues are to be addressed:

1. Choosing the factorial portion of the design
2. Number of center points
3. Determining the $\alpha$ value for the axial points

***Choice of factorial portion of the design:*** A CCD should have the total number of distinct design points $N = n_F + 2k + 1$, must be at least $\dfrac{(k+1)(k+2)}{2}$

where,

$k$ = number of factors,
$n_F$ = number of factorial points,
$2k$ = number of star points

$\dfrac{(k+1)(k+2)}{2}$ = smallest number of points required for the second order model to be estimated.

Table 8.1 can be used to select a CCD with economical run size for $2 \le k \le 4$.

**TABLE 8.1** Selection of CCD with economical run size

| $k$ | $(k + 1)(k + 2)/2$ | $N$ | $n_F$ | *Factorial portion (cube points)* |
|---|---|---|---|---|
| 2 | 6 | 7 | 2 | $2^{2-1}(I = AB)$ |
| 2 | 6 | 9 | 4 | $2^2$ |
| 3 | 10 | 11 | 4 | $2_{\text{III}}^{3-1}(I = ABC)$ |
| 3 | 10 | 15 | 8 | $2^3$ |
| 4 | 15 | 17 | 8 | $2_{\text{III}}^{4-1}(I = ABD)$ |
| 4 | 15 | 20 | 11 | $11 \times 4$ sub matrix of 12 run PB design* |
| 4 | 15 | 25 | 16 | $2^4$ |

*Obtained by taking the first four columns and deleting row 3 of the 12-run Plackett–Burman design.

Further discussion on small composite designs on the Placket–Burman designs can be found in Draper and Lin (1990).

### *Choice of $\alpha$ value*

In general $\alpha$ value is selected between 1 and $\sqrt{k}$. A design that produces contours of constant standard deviation of predicted response is called *rotatable design*. A CCD is made rotatable by selecting

$$\alpha = (n_F)^{1/4} \tag{8.4}$$

This design provides good predictions throughout the region of interest. If the region of interest is spherical, the best choice of $\alpha = \sqrt{k}$. This design is often called *spherical* CCD.

In general, the choice of $\alpha$ depends on the geometric nature of and practical constraints on the design region. Unless some practical considerations dictate the choice of $\alpha$, Eq. (8.4) serves as a useful guideline.

### *Number of runs at the center point*

The rule of thumb is to have three to five center points when $\alpha$ is close to $\sqrt{k}$. If the purpose is to obtain the error it is better to have more than 4 or 5 runs. Further discussion on choice of $\alpha$ and the number of runs at the center point can be found in Box and Draper (1987). The graphical representation of CCD for 2 factors is shown in Figure 8.2. The CCD for $k$ = 2 and $k$ = 3 with five centre points is given in Tables 8.2 and 8.3 respectively.

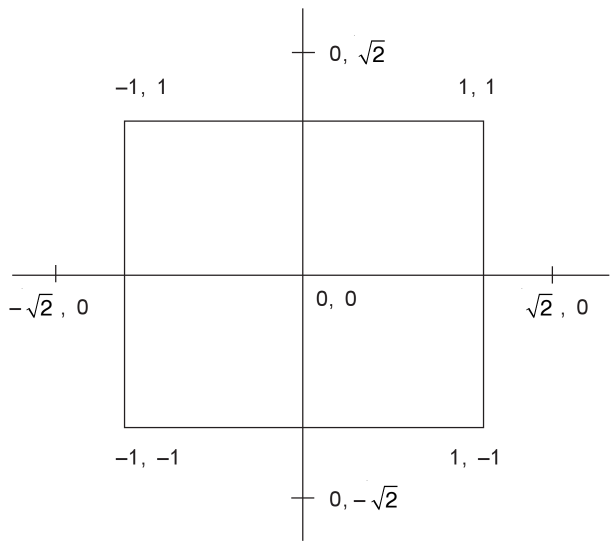



**FIGURE 8.2** The central composite design for two factors.

**TABLE 8.2** Central composite design for $k = 2$

| *Run* | *Factors* | |
|---|---|---|
| | *A* | *B* |
| 1 | –1 | –1 |
| 2 | 1 | –1 |
| 3 | –1 | 1 |
| 4 | 1 | 1 |
| 5 | –1.41 | 0 |
| 6 | 1.41 | 0 |
| 7 | 0 | –1.14 |
| 8 | 0 | 1.14 |
| 9 | 0 | 0 |
| 10 | 0 | 0 |
| 11 | 0 | 0 |
| 12 | 0 | 0 |
| 13 | 0 | 0 |

($\alpha = \sqrt{k}$ and $n_c = 5$)

**TABLE 8.3** Central composite design for $k = 3$

| *Run* | *Factors* | | |
|---|---|---|---|
| | *A* | *B* | *C* |
| 1 | –1 | –1 | –1 |
| 2 | 1 | –1 | –1 |
| 3 | –1 | 1 | –1 |
| 4 | 1 | 1 | –1 |
| 5 | –1 | –1 | 1 |
| 6 | 1 | –1 | 1 |
| 7 | –1 | 1 | 1 |
| 8 | 1 | 1 | 1 |
| 9 | –1.73 | 0 | 0 |
| 10 | 1.73 | 0 | 0 |
| 11 | 0 | –1.73 | 0 |
| 12 | 0 | 1.73 | 0 |
| 13 | 0 | 0 | –1.73 |
| 14 | 0 | 0 | 1.73 |
| 15 | 0 | 0 | 0 |
| 16 | 0 | 0 | 0 |
| 17 | 0 | 0 | 0 |
| 18 | 0 | 0 | 0 |
| 19 | 0 | 0 | 0 |

($\alpha = \sqrt{k}$ and $n_c = 5$)

### 8.2.3 Box–Behnken Designs

Box and Behnken (1960) developed a set of three-level second-order response surface designs. These were developed by combining two-level factorial designs with balanced or partially balanced incomplete block designs. The construction of the smallest design for three factors is explained below. A balanced incomplete block design with three treatments and three blocks is given by

| *Block* | *Treatments* | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| 1 | X | X | |
| 2 | X | | X |
| 3 | | X | X |

The three treatments are considered as three factors $A$, $B$, $C$ in the experiment. The two crosses (Xs) in each block are replaced by the two columns of the $2^2$ factorial design and a column of zeros are inserted in the places where the cross do not appear. This procedure is repeated for the next two blocks and some centre points are added resulting in the Box-Behnken design for $k = 3$ which is given in Table 8.4. Similarly, designs for $k = 4$ and $k = 5$ can be obtained. The advantage of this design is that each factor requires only three levels.

**TABLE 8.4** Box–Behnken design for $k = 3$

| *Run* | *A* | *B* | *C* |
|---|---|---|---|
| 1 | –1 | –1 | 0 |
| 2 | –1 | 1 | 0 |
| 3 | 1 | –1 | 0 |
| 4 | 1 | 1 | 0 |
| 5 | –1 | 0 | –1 |
| 6 | –1 | 0 | 1 |
| 7 | 1 | 0 | –1 |
| 8 | 1 | 0 | 1 |
| 9 | 0 | –1 | –1 |
| 10 | 0 | –1 | 1 |
| 11 | 0 | 1 | –1 |
| 12 | 0 | 1 | 1 |
| 13 | 0 | 0 | 0 |
| 14 | 0 | 0 | 0 |
| 15 | 0 | 0 | 0 |

## 8.3 ANALYSIS OF DATA FROM RSM DESIGNS

In this section we will discuss how polynomial models can be developed using $2^k$ factorial designs.

Fitting of a polynomial model can be treated as a particular case of multiple linear regression. The experimental designs used to develop the regression models must facilitate easy formation of least squares normal equations so that solution can be obtained with least difficulty.

### 8.3.1 Analysis of First-order Design

Let us consider a $2^3$ factorial design to illustrate a general approach to fit a linear equation. Suppose we have the following three factors studied each at two levels in an experiment.

| *Factors* | *Level* | |
|---|---|---|
| | *Low* (–1) | *High* (+) |
| Time (min) (*A*) | 6 | 8 |
| Temperature (*B*) | 240°C | 300°C |
| Concentration (%) (*C*) | 10 | 14 |

Let the first-order model to be fitted is

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \tag{8.5}$$

where $X_1$, $X_2$, $X_3$ are the coded variables with ±1 as high and low levels of the factors *A*, *B* and *C* respectively.

The relation between the coded variable ($X$) and the natural variables is given by

$$\frac{V - (h + l)/2}{(h - l)/2} \tag{8.6}$$

$V$ = natural/actual variable such as time, temperature etc.
$h$ = value of high level of the factor
$l$ = value of low level of the factor

Accordingly, we have

$$X_1 = \frac{\text{Time} - 7}{1.0}, X_2 = \frac{\text{Temperature} - 270}{30}, X_3 = \frac{\text{Concentration} - 12}{2.0}$$

Suppose we have the following experimental design and data (Table 8.5).

**TABLE 8.5** Data for fitting first-order model

| *Run* | *Factors* | | | *Y* |
|---|---|---|---|---|
| | $X_1$ | $X_2$ | $X_3$ | |
| (1) | –1 | –1 | –1 | 61 |
| *a* | 1 | –1 | –1 | 83 |
| *b* | –1 | 1 | –1 | 51 |
| *ab* | 1 | 1 | –1 | 70 |
| *c* | –1 | –1 | 1 | 66 |
| *ac* | 1 | –1 | 1 | 92 |
| *bc* | –1 | 1 | 1 | 56 |
| *abc* | 1 | 1 | 1 | 83 |

The data is analysed as follows. In order to facilitate easy computation of model parameters using least-square method, a column $X_0$ with +1s are added to the design in Table 8.5 as given in Table 8.6.

**TABLE 8.6** First-order model with three variables

| $X_0$ | $X_1$ | $X_2$ | $X_3$ | $y$ |
|---|---|---|---|---|
| 1 | –1 | –1 | –1 | 61 |
| 1 | 1 | –1 | –1 | 83 |
| 1 | –1 | 1 | –1 | 51 |
| 1 | 1 | 1 | –1 | 70 |
| 1 | –1 | –1 | 1 | 66 |
| 1 | 1 | –1 | 1 | 92 |
| 1 | –1 | 1 | 1 | 56 |
| 1 | 1 | 1 | 1 | 83 |

Obtain sum of product of each column with response $y$.

$$0y = \Sigma X_0\, y = 562 \text{ (Grand total)}$$
$$1y = \Sigma X_1\, y = 94$$
$$2y = \Sigma X_2\, y = -42$$
$$3y = \Sigma X_3\, y = 32$$

This design (Table 8.5) is simply a $2^3$ design. Hence these sums are nothing but contrast totals.

That is, $C_A = 94$, $C_B = -42$, $C_C = 32$

Therefore, the regression coefficients can be estimated as follows:

$$\beta_0 = \bar{y}$$
$$= \frac{562}{8} = 70.25$$

$$\beta_1 = \frac{1}{2} \text{ (Effect of } A) = \frac{1}{2}\left(\frac{C_A}{n2^{k-1}}\right)$$

where,

$\bar{y}$ = grand mean

$C_A$ = contrast $A$ = 94

$n$ = number of replications = 1

$\beta_1$ = 94/8 = 11.75

Similarly, $\beta_2 = -42/8 = -5.25$ and $\beta_3 = 32/8 = 4.00$

The fitted regression model is

$$\hat{Y} = 70.25 + 11.75\, X_1 - 5.25\, X_2 + 4.0\, X_3 \tag{8.7}$$

***Analysis of variance:*** Usually, the total sum of squares is partitioned into SS due to regression (linear model) and residual sum of squares. And

SS residual ($SS_R$) = Experimental error SS or pure SS + SS lack of fit

In this illustration, the experiment is not replicated and hence pure SS is zero. Therefore,

$$SS_{\text{Total}} = SS_{\text{Linear model}} + SS_{\text{Lack of fit}} \tag{8.8}$$

These sum of squares are computed as follow:

$$SS_{\text{Total}} = \sum_{i=1}^{N} y_i^2 - \text{CF}$$

$$= \{(61)^2 + (83)^2 + \cdots + (83)^2\} - \frac{(562)^2}{8}$$

$$= 40956.0 - 39480.5 = 1475.50$$

$SS_{\text{Model}}$ = All factor sum of squares. This can be computed from the contrast totals, which is given below:

$$SS_A = \frac{(C_A)^2}{n2^k} = \frac{(94)^2}{8} = 1104.5 \quad \text{(Since } n = 1 \text{ and } k = 3\text{)}$$

Similarly, $$SS_B = \frac{(-42)^2}{8} = 220.5 \text{ and } SS_C = \frac{(32)^2}{8} = 128.0$$

$$SS_{\text{Linear model}} = 1104.5 + 220.5 + 128.0 = 1453$$

The ANOVA for the first-order model is given in Table 8.7.

**TABLE 8.7** ANOVA for the first-order model

| *Source* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ |
|---|---|---|---|---|
| Linear model | 1453.00 | 3 | 484.33 | 86.02 |
| Lack of fit (residual) | 22.50 | 4 | 5.63 | |
| Total | 1475.50 | 7 | | |

When tested with the lack of fit, the linear model is a good fit to the data.

Note that SS lack of fit = $SS_{AB} + SS_{AC} + SS_{BC} + SS_{ABC.}$

As these types of designs do not provide any estimate of experimental error variance, the lack of fit cannot be tested. A significant lack of fit indicates the model inadequacy. In practice we need experimental error when first-order or second-order designs are used. Otherwise the experimenter does not know whether the equation adequately represents the true surface. To obtain the experimental error either the whole experiment is replicated or some center points be added to the $2^k$ factorial design. Addition of center points do not alter the estimate of the regression coefficients except that $\beta_0$ becomes the mean of the whole experiment. Suppose $n_c$ number of center points are added. These observations are used to measure experimental error and will have $n_c - 1$ degrees of freedom. If $\overline{Y}_C$ is the mean of these observations and $\overline{Y}_F$ is the mean of factorial observations, $(\overline{Y}_F - \overline{Y}_C)$ gives an additional degree of freedom for measuring the lack of fit. The sum of squares for lack of fit is given by

$$\frac{n_c n_F}{n_c + n_F} (\overline{Y}_F - \overline{Y}_C)^2 \tag{8.9}$$

where $n_F$ is the number of factorial points.

The effects are estimated using contrasts. The effects and coefficients for the first-order model are summarized in Table 8.8.

**TABLE 8.8** Effects and coefficients for the first-order model

| *Term in the model* | *Effect estimate* | *Coefficient* |
|---|---|---|
| Constant | | 70.25 |
| Time (*A*) | 23.50 | 11.75 |
| Temperature (*B*) | –10.50 | –5.25 |
| Concentration (*C*) | 8.00 | 4.00 |
| *AB* | – 0.50 | – 0.25 |
| *AC* | 3.00 | 1.50 |
| *BC* | 1.00 | 0.50 |
| *ABC* | 1.00 | 0.50 |

Figures 8.3 and 8.4 show the normal and half-normal plot of the effects respectively for the first-order design given in Table 8.5. From the regression model [Eq. (8.7)], it is evident that high value of time and concentration and low value of temperature produces high response. This is also evident from the contour plot of response (Figure 8.5). Because the fitted model is first order (includes only main effects), the surface plot of the response is a plane (Figure 8.6). The response surface can be analysed using computer software to determine the optimum condition.

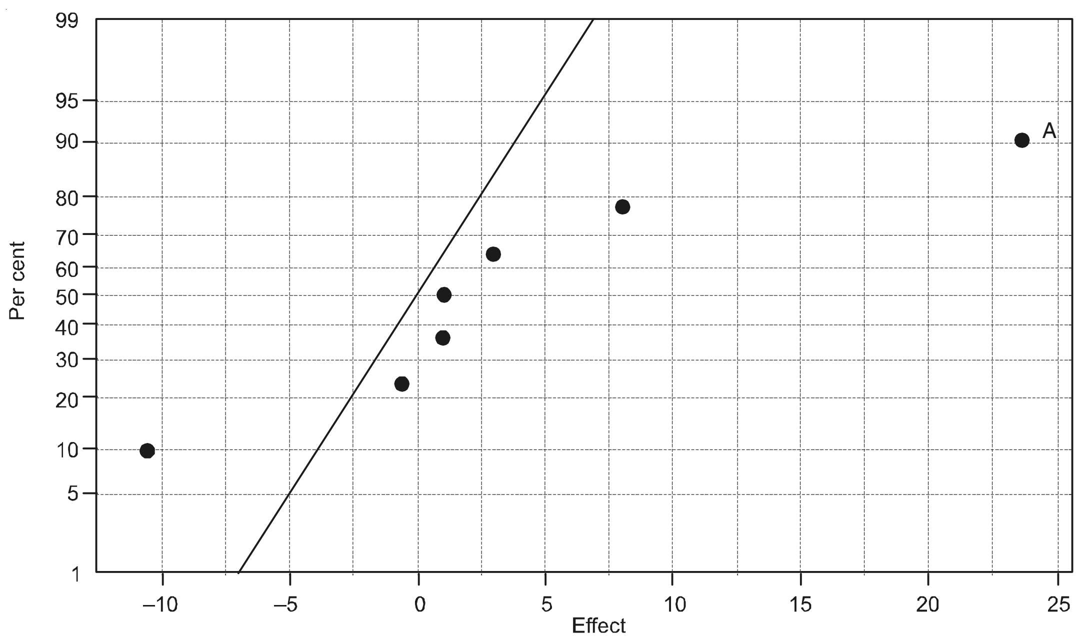



**FIGURE 8.3** Normal plot of the effects for the first order design given in Table 8.5.

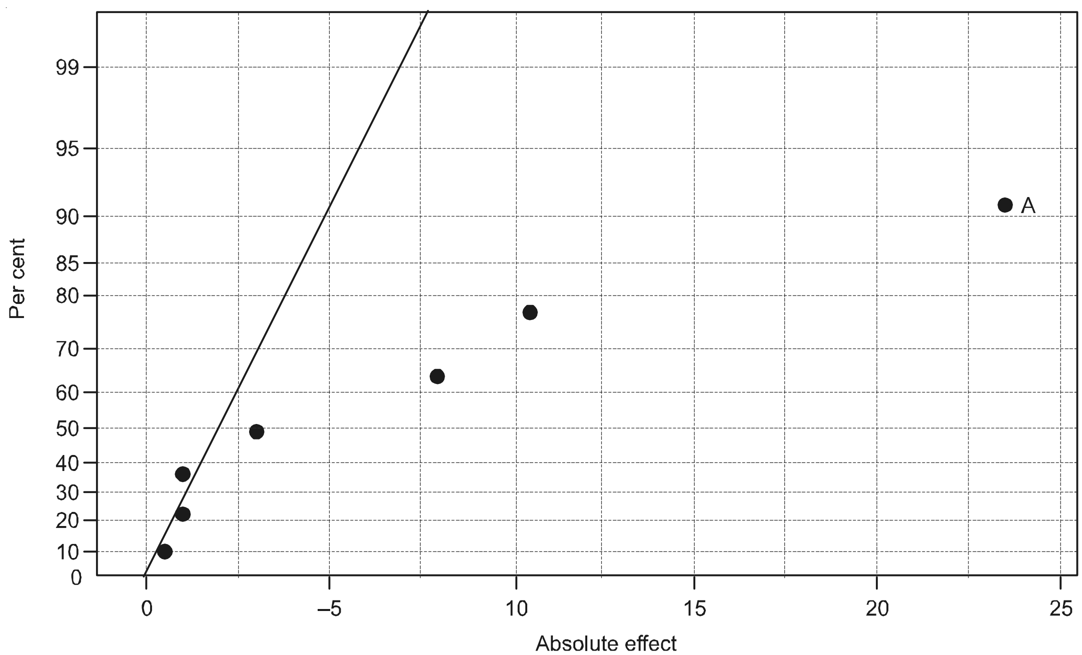



**FIGURE 8.4** Half-normal plot of the effects for the first-order design given in Table 8.5.

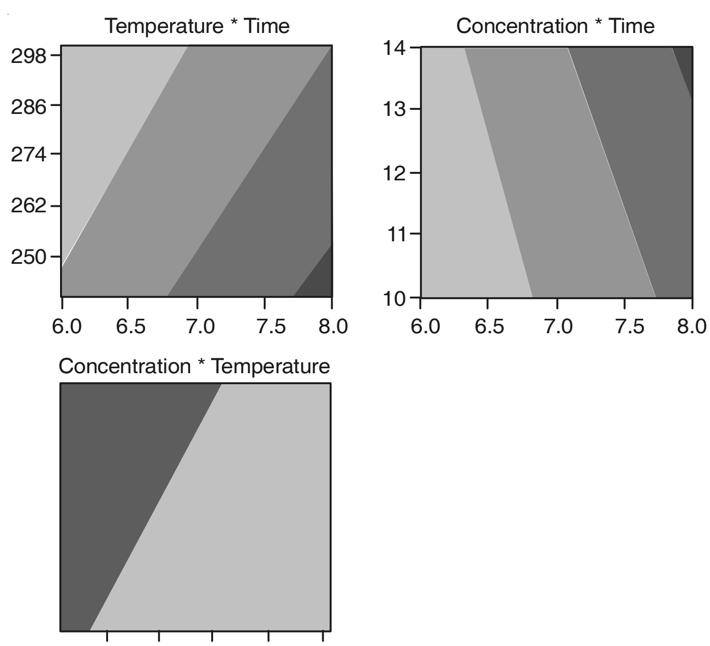



**FIGURE 8.5** Contour plots for the first-order design given in Table 8.5.

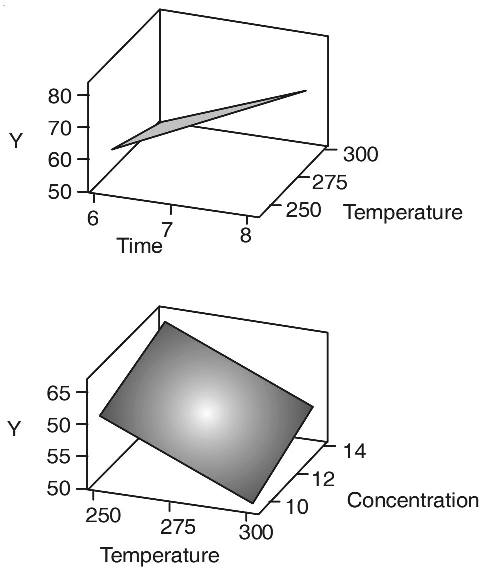



**FIGURE 8.6** Surface plots of response for the first order design given in Table 8.5.

### 8.3.2 Analysis of Second-order Design

The general form of a second-order polynomial with two variables is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 + e \tag{8.10}$$

In order to estimate the regression coefficients, we normally use CCD with two factors (Table 8.2). This design has $k = 2$, $n_F = 4$, $n_C = 5$ and $\alpha = \sqrt{2}$. The CCD for two factors with data is given in Table 8.9.

For conducting the experiment, the coded variables ($X_i$) are converted into natural variables (temperature, time, pressure etc.) using the following relation:

$$X_i = \frac{\xi - (A_H + A_L)/2}{(A_H - A_L)/2} \tag{8.11}$$

where,

$\xi$ = natural variable
$A_H$ = value of high level of factor $A$
$A_L$ = value of low level of factor $A$

Suppose, the factor $A$ is the temperature with low level = 240° and high level = 300°.

Then,
$$X_i = \frac{\text{Temperature} - 270}{30} \tag{8.12}$$

Now when $X_i = 0$ (center point), the level of the factor is set at 270° [from Eq. (8.12)] Similarly, when $X_i = 1.682$ (the axial point), the factor level will be about 320°.

Following this procedure, the experimental design can be converted into an experimental layout.

### *Fitting the model*

*Step* 1: To the CCD (Table 8.2) add column $X_0$ to the left and $X_1^2$, $X_2^2$, $X_1X_2$ and $y$ (response) to the right side as given in Table 8.9.

*Step* 2: Obtain sum of product of each column with $y$. Denote these sum of products as ($0y$) for the $X_0$ column, $1y$ for $X_1$ column and so on. That is,

$$0y = \Sigma X_0\, y,\ 1y = \Sigma X_1\, y,\ 2y = \Sigma X_2\, y,\ 11y = \Sigma X_1^2\, y,\ 22y = \Sigma X_2^2\, y \text{ and } 12y = \Sigma X_1 X_2\, y$$

*Step* 3: Obtain regression coefficients using the following formulae:

$$\beta_0 = 0.2\,(0y) - 0.10\,\Sigma\, iiy$$

where,

$\Sigma iiy = 11y + 22y$

$\beta_i = 0.125\ (iy)$

$\beta_{ii} = 0.125\ (iiy) + 0.01875\ \Sigma iiy - 0.10\ (0y)$

$\beta_{ij} = 0.25\ (ijy)$

The quantity $\Sigma iiy$ is sum of cross products of all the squared terms with $y$.

In this model $\Sigma iiy$ is $\Sigma X_1^2\, y + \Sigma X_2^2\, y$

For the data shown, the parameters of the model are estimated as follows:

$$0y = \Sigma X_0\, y = 110.20$$
$$1y = \Sigma X_1\, y = 7.96$$
$$2y = \Sigma X_2\, y = 4.12$$
$$11y = \Sigma X_1^2\, y = 59$$
$$22y = \Sigma X_2^2\, y = 62$$
$$12y = \Sigma X_1 X_2\, y = 1.00$$

**TABLE 8.9** Data for central composite design with two factors

| $X_0$ | $X_1$ | $X_2$ | $X_1^2$ | $X_2^2$ | $X_1X_2$ | $y$ |
|---|---|---|---|---|---|---|
| 1 | –1 | –1 | 1 | 1 | 1 | 6.5 |
| 1 | 1 | –1 | 1 | 1 | –1 | 8.0 |
| 1 | –1 | 1 | 1 | 1 | –1 | 7.0 |
| 1 | 1 | 1 | 1 | 1 | 1 | 9.5 |
| 1 | –1.414 | 0 | 2 | 0 | 0 | 5.6 |
| 1 | 1.414 | 0 | 2 | 0 | 0 | 8.4 |
| 1 | 0 | –1.414 | 0 | 2 | 0 | 7.0 |
| 1 | 0 | 1.414 | 0 | 2 | 0 | 8.5 |
| 1 | 0 | 0 | 0 | 0 | 0 | 9.9 |
| 1 | 0 | 0 | 0 | 0 | 0 | 10.3 |
| 1 | 0 | 0 | 0 | 0 | 0 | 10.0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 9.7 |
| 1 | 0 | 0 | 0 | 0 | 0 | 9.8 |

The regression coefficients are estimated as follows:

$$\beta_0 = 0.2\ (0y) - 0.10\ \Sigma\ iiy = 0.2\ (110.20) - 0.10\ (59 + 62) = 9.94$$

$$\beta_1 = 0.125\ (1y) = 0.125\ (7.96) = 0.995$$

$$\beta_2 = 0.125\ (2y) = 0.125\ (4.12) = 0.515$$

$$\beta_{11} = 0.125\ (iiy) + 0.01875\ \Sigma\ iiy - 0.10\ (0y)$$

$$= 0.125\ (59) + 0.01875\ (121) - 0.10\ (110.2) = -1.38$$

$$\beta_{22} = 0.125\ (62) + 0.01875\ (121) - 0.10\ (110.2) = -1.00$$

$$\beta_{12} = 0.25\ (ijy)$$

$$= 0.25\ (1.00) = 0.25$$

Therefore, the fitted model is

$$\hat{Y} = 9.94 + 0.995\ X_1 + 0.515\ X_2 - 1.38\ X_1^2 - 1.00\ X_2^2 + 0.25\ X_1X_2 \tag{8.13}$$

Equation (8.13) can be used to study the response surface. Usually, we use computer software for studying these designs. The software gives the model analysis including all the statistics and the relevant plots. Figure 8.7 and 8.8 show the contour plot and surface plot of response respectively for the Central Composite Design (CCD) with two factors *A* and *B* discussed above.

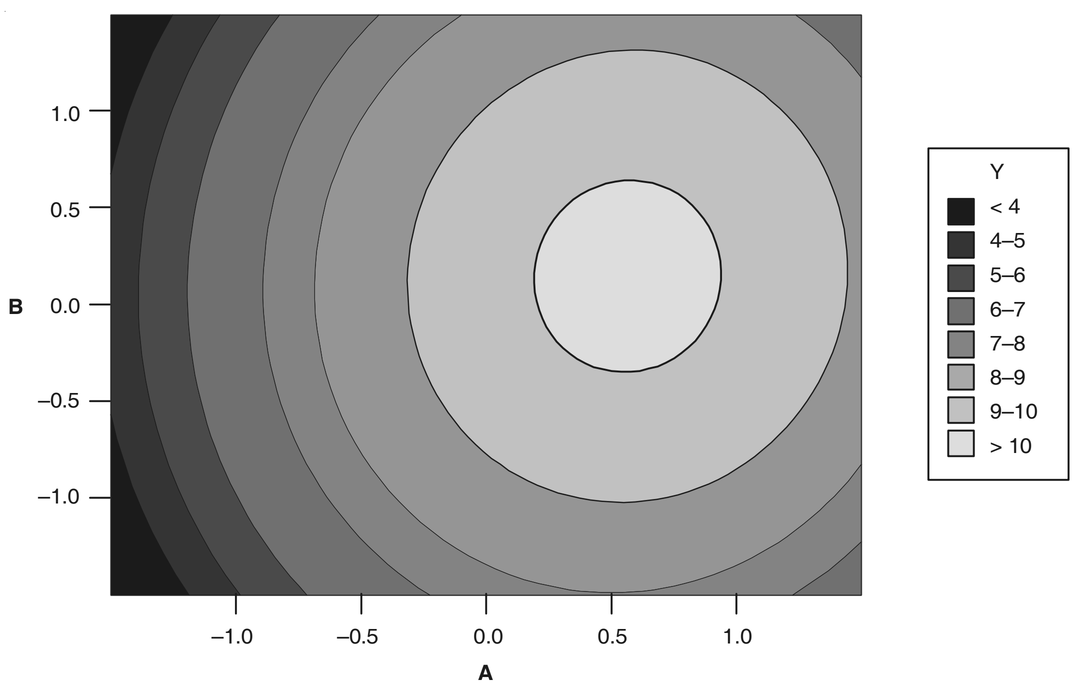



**FIGURE 8.7** Contour plot of response for CCD with two factors.

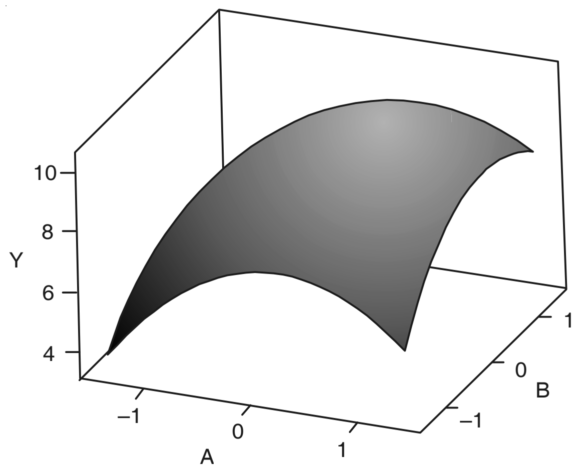



**FIGURE 8.8** Surface plot of response for CCD with two factors.

### *CCD with three variables*

Central Composite Design with three variables (factors) is given in Table 8.10. Usually, we use computer software for solution of these models which will give the fitted model including ANOVA and other related statistics and generate contour and response surface plots. Note that $y$ in Table 8.10 is the response column to be obtained from the experiment.

**TABLE 8.10** Data for CCD with three factors

| $X_0$ | $X_1$ | $X_2$ | $X_3$ | $X_1^2$ | $X_2^2$ | $X_3^2$ | $X_1X_2$ | $X_1X_3$ | $X_2X_3$ | $y$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | –1 | –1 | –1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| 1 | 1 | –1 | –1 | 1 | 1 | 1 | –1 | –1 | 1 | |
| 1 | –1 | 1 | –1 | 1 | 1 | 1 | –1 | 1 | –1 | |
| 1 | 1 | 1 | –1 | 1 | 1 | 1 | 1 | –1 | –1 | |
| 1 | –1 | –1 | 1 | 1 | 1 | 1 | 1 | –1 | –1 | |
| 1 | 1 | –1 | 1 | 1 | 1 | 1 | –1 | 1 | –1 | |
| 1 | –1 | 1 | 1 | 1 | 1 | 1 | –1 | –1 | 1 | |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| 1 | –1.682 | 0 | 0 | 2.828 | 0 | 0 | 0 | 0 | 0 | |
| 1 | 1.682 | 0 | 0 | 2.828 | 0 | 0 | 0 | 0 | 0 | |
| 1 | 0 | –1.682 | 0 | 0 | 2.828 | 0 | 0 | 0 | 0 | |
| 1 | 0 | 1.682 | 0 | 0 | 2.828 | 0 | 0 | 0 | 0 | |
| 1 | 0 | 0 | –1.682 | 0 | 0 | 2.828 | 0 | 0 | 0 | |
| 1 | 0 | 0 | 1.682 | 0 | 0 | 2.828 | 0 | 0 | 0 | |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |

## PROBLEMS

**8.1** Use the data from Table 8.11 $2^3$ design and fit a first-order model.

**TABLE 8.11** Data for Problem 8.1

| Factors | | | |
|---|---|---|---|
| $X_1$ | $X_2$ | $X_3$ | $Y$ |
| –1 | –1 | –1 | 31 |
| 1 | –1 | –1 | 43 |
| –1 | 1 | –1 | 34 |
| 1 | 1 | –1 | 47 |
| –1 | –1 | 1 | 45 |
| 1 | –1 | 1 | 37 |
| –1 | 1 | 1 | 50 |
| 1 | 1 | 1 | 41 |

**8.2** The data given in Table 8.12 have been collected from a Central Composite Design with two factors. Fit a second-order model.

**TABLE 8.12** Data for Problem 8.2

| $X_1$ | $X_2$ | $y$ |
|---|---|---|
| –1 | –1 | 34 |
| 1 | –1 | 26 |
| –1 | 1 | 13 |
| 1 | 1 | 26 |
| –1.414 | 0 | 30 |
| 1.414 | 0 | 31 |
| 0 | –1.414 | 18 |
| 0 | 1.414 | 30 |
| 0 | 0 | 20 |
| 0 | 0 | 18 |
| 0 | 0 | 23 |
| 0 | 0 | 22 |
| 0 | 0 | 20 |

# PART II

# Taguchi Methods

# CHAPTER 9 Quality Loss Function

## 9.1 INTRODUCTION

Quality has been defined in different ways by different experts over a period of time. Some of them are as follows:

- Fitness for use
- Conformance to specifications
- Customer satisfaction/delight
- The totality of features and characteristics that satisfy the stated and implied needs of the customer

The simplest definition of high quality is a happy customer. The true evaluator of quality is the customer. The quality of a product is measured in terms of the features and characteristics that describe the performance relative to customer requirements or expectations. To satisfy the customer, the product must be delivered in right quality at right time at right place and provide right functions for the right period at right price.

## 9.2 TAGUCHI DEFINITION OF QUALITY

The quality of a product is defined as the loss imparted by the product to society from the time the product is shipped to the customer. The loss may be due to failure, repair, variation in performance, pollution, noise, etc. A truly high quality product will have a minimal loss to society.

The following are the types of loss:

- Product returns
- Warranty costs
- Customer complaints and dissatisfaction
- Time and money spent by the customer
- Eventual loss of market share and growth

Under warranty, the loss will be borne by the manufacturer and after warranty it is to be paid by the customer. Whoever pays the loss, ultimately it is the loss to the society.

## 9.3 TAGUCHI QUALITY LOSS FUNCTION

The loss which we are talking about is the loss due to functional variation/process variation. Taguchi quantified this loss through a quality loss function. The quality characteristic is the object of interest of a product or process. Generally, the quality characteristic will have a target. There are three types of targets.

***Nominal–the best:*** When we have a characteristic with bi-lateral tolerance, the nominal value is the target. That is, if all parts are made to this value, the variation will be zero and it is the best.

*For example:* A component with a specification of 10 ± 0.01 mm has the nominal value of 10 mm. Similarly, if the supply voltage has a specification of 230 ± 10 V. Here the nominal value is 230 V.

***Smaller–the better:*** It is a non negative measurable characteristic having an ideal target as zero.

*For example:* Tyre wear, pollution, process defectives, etc.

***Larger–the better:*** It is also a non negative measurable characteristic that has an ideal target as infinity (∞).

*For example:* Fuel efficiency, strength values, etc.

For each quality characteristic, there exist some function which uniquely defines the relation between economic loss and the deviation of the quality characteristic from its target. Taguchi defined this relation as a quadratic function termed Quality Loss Function (QLF).

### 9.3.1 Quality Loss Function (Nominal—the best case)

When the quality characteristic is of the type Nominal–the best, the quality loss function is given by

$$L(Y) = K(Y - T)^2 \tag{9.1}$$

where,

$Y$ = value of the quality characteristic (e.g., length, force, diameter etc.)

$L(Y)$ = loss in ₹ per product when the quality characteristic is equal to $Y$

$T$ = target value of $Y$

$K$ = proportionality constant or cost coefficient which depends on the cost at the specification limits and the width of the specification

Figure 9.1 shows the QLF for nominal—the best case.

When $Y = T$, the loss is '0'. From Figure 9.1, it can be observed that, as $Y$ deviates from the target $T$, the quality loss increases on either side of $T$. Note that a loss of $A_0$ is incurred even at the specification limit or consumer tolerance ($\Delta_0$). This loss is attributed to the variation in quality of performance (functional variation). Thus, confirmation to specification is an inadequate measure of quality. The quality loss is due to customer dissatisfaction. It can be related to the product quality characteristics. And QLF is a better method of assessing quality loss compared to the traditional method which is based on proportion of defectives.

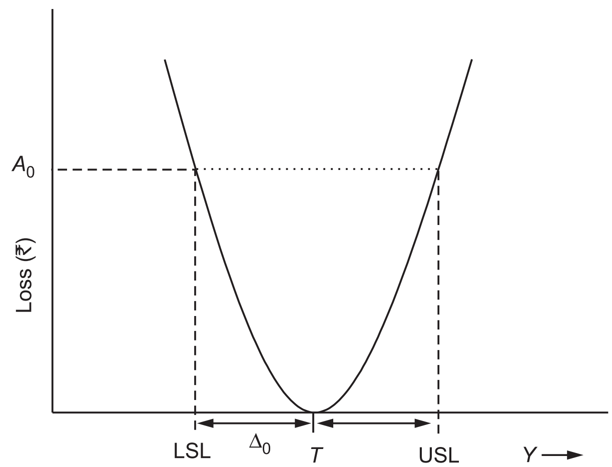



**FIGURE 9.1** QLF for nominal—the best case.

Suppose the supply voltage for a compressor is the quality characteristic ($Y$). The specification for voltage is 230 ± 10 V.

Target voltage = 230 V

Consumer tolerance ($\Delta_0$) = ±10 V.

Let $A_0$ = average cost of repair/adjustment at a deviation of $\Delta_0$

$$L(Y) = K(Y - T)^2$$

$$K = \frac{L(Y)}{(Y - T)^2}$$

At the specification,
$$K = \frac{A_0}{\Delta_0^2} \tag{9.2}$$

Therefore,
$$L(Y) = \frac{A_0}{\Delta_0^2}(Y - T)^2 \tag{9.3}$$

If $A_0$ = ₹ 500 at a deviation of 10 V, then the cost coefficient

$$K = \frac{500}{(10\text{ V})^2} = ₹\ 5/\text{V}^2$$

And $L(Y) = 5.00(Y - 230)^2$

At $Y = T(230)$, the target value, the loss is zero.

For a 5 V deviation,

$$L(235) = 5.00(235 - 230)^2 = ₹\ 125.00$$

Thus, we can estimate the loss for any amount of deviation from the target.

***Estimation of average quality loss:*** The quality loss we obtain from Eq. (9.1) is for a single product. The quality loss estimate should be an average value and should be estimated from a sample of parts/products. Suppose we take a sample of $n$ parts/products. Then we will have

$$Y_1, Y_2, ..., Y_i \text{ values}$$

where, $i = 1, 2, 3, ..., n$

Let the average of this sample = $\overline{Y}$. This $\overline{Y}$ may or may not be equal to the target value ($T$).

The quality loss is modelled as follows:

$$L(Y) = K(Y - T)^2$$

$$= \frac{K(Y_1 - T)^2 + K(Y_2 - T)^2 + \cdots + K(Y_n - T)^2}{n} \tag{9.4}$$

$$= \frac{K[(Y_1 - T)^2 + (Y_2 - T)^2 + \cdots + (Y_n - T)^2]}{n} \tag{9.5}$$

The term in the parenthesis is the average of all values of $(Y_i - T)^2$ and is called *Mean Square Deviation* (MSD).

Therefore, $L(Y) = K(\text{MSD})$

$$\text{MSD} = \frac{[(Y_1 - T)^2 + (Y_2 - T)^2 + \cdots + (Y_n - T)^2]}{n} \tag{9.6}$$

$$= \frac{1}{n} \Sigma(Y_i - T)^2$$

On simplification, $$\text{MSD} = \sigma_n^2 + (\overline{Y} - T)^2 \tag{9.7}$$

where,

$\sigma_n$ = population standard deviation

$(\overline{Y} - T)^2$ = bias of the sample from the target.

For practical purpose we use sample standard deviation ($\sigma_{n-1}$)

Since, $L(Y) = K(\text{MSD})$

$$L(Y) = K[\sigma^2 + (\overline{Y} - T)^2] \tag{9.8}$$

Note that the loss estimated by Eq. (9.8) is the average loss per part/product ($\rho$)

$$\text{Total loss} = \rho(\text{number of parts/products})$$

### 9.3.2 Quality Loss Function (Smaller—the better case)

For this case the target value $T = 0$

Substituting this in Eq. (9.1), we get

$$L(Y) = K(Y - 0)^2$$

$$= K\,Y^2 \tag{9.9}$$

and $$K = \frac{L(Y)}{Y^2} = \frac{A_0}{\Delta_0^2} \tag{9.10}$$

Figure 9.2 shows the loss function for smaller—the better type of quality characteristic.

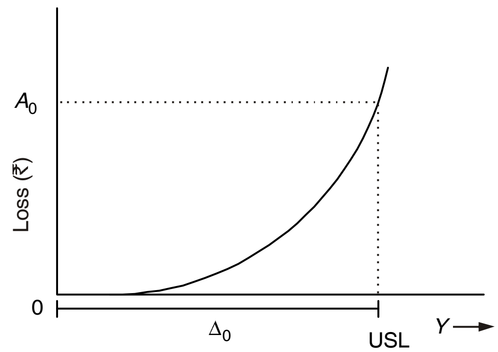



**FIGURE 9.2** QLF for smaller—the better case.

The average loss per part/product can be obtained by substituting $T = 0$ in Eq. (9.8).

$$\begin{aligned} L(Y) &= K[\sigma^2 + (\bar{Y} - T)^2] \\ &= K[\sigma^2 + (\bar{Y} - 0)^2] \\ &= K(\sigma^2 + \bar{Y}^2) \end{aligned} \tag{9.11}$$

### 9.3.3 Quality Loss Function (Larger—the better case)

Mathematically, a larger–the better characteristic can be considered as the inverse of a smaller–the better characteristic. Therefore, we can obtain $L(Y)$ for this case from Eq. (9.9). That is

$$L(Y) = K\left(\frac{1}{Y^2}\right) \tag{9.12}$$

and $$\begin{aligned} K &= L(Y)\ Y^2 \\ &= A_0\Delta_0^2 \end{aligned} \tag{9.13}$$

Figure 9.3 shows the quality loss for the case of larger–the better type of quality characteristic. The loss per part/product averaged from many sample parts/products is given by

$$L(Y) = K \times \frac{1}{n}\left(\frac{1}{Y_1^2} + \frac{1}{Y_2^2} + \cdots + \frac{1}{Y_n^2}\right)$$

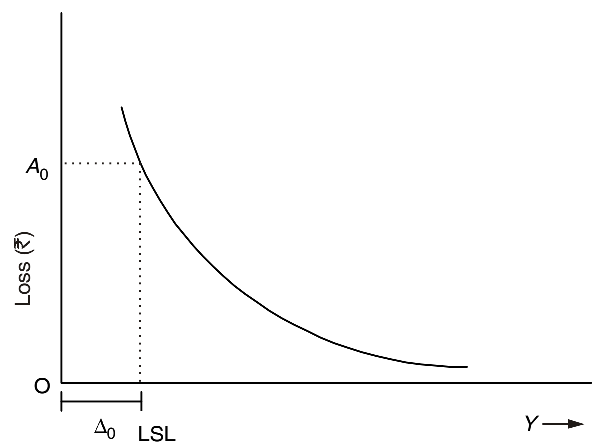



**FIGURE 9.3** QLF for larger—the better type of quality characteristic.

$$= K \times \frac{1}{n} \Sigma\left(\frac{1}{Y_i^2}\right) \tag{9.14}$$

$$= K(\text{MSD})$$

On simplification of Eq. (9.14), the average loss per part/product will be given by

$$L(Y) = \frac{K}{\overline{Y}^2}\left[1 + \frac{3\sigma^2}{\overline{Y}^2}\right] \tag{9.15}$$

## ILLUSTRATION 9.1

**Nominal–the Best Case**

A company manufactures coolers of various capacities in which the compressor forms the critical item. The specification for the voltage of the compressor of a specific capacity is 230 ± 20 V. The consumer loss at the specification was estimated to be ₹ 3000. This includes repair/replacement cost and other related cost based on customer complaints. The company purchases two types of compressors (Brand *A* and Brand *B*). The output voltage data collected for 20 compressors are given in Table 9.1.

**TABLE 9.1** Data for Illustration 9.1 (compressor voltage)

| | |
|---|---|
| Brand *A* | 230, 245, 230, 210, 212, 215, 240, 255, 240, 225, 240, 215, 255, 230, 212, 230, 240, 245, 225, 250 |
| Brand *B* | 235, 240, 230, 245, 235, 245, 230, 250, 230, 250, 235, 240, 250, 230, 245, 235, 245, 230, 250, 230 |

Determine which brand of compressors should be purchased.

**SOLUTION:**

Target value ($T$) = 230 V
Consumer tolerance ($\Delta_0$) = ± 20 V
Consumer loss ($A_0$) = ₹ 3000

$$L(Y) = K(Y - T)^2$$
$$3000 = K(20)^2$$
$$K = 7.5$$
$$L(Y) = K(\text{MSD})$$
$$= 7.5[\sigma^2 + (\bar{Y} - T)^2]$$

where, $\bar{Y} = \dfrac{\Sigma Y_i}{n}$ and $\sigma^2 = \dfrac{\Sigma Y_i^2 - n\bar{Y}^2}{n - 1}$, $n = 20$ and $T = 230$

The mean and standard deviation values computed for the two brands of compressors and the average loss per compressor are given in Table 9.2.

**TABLE 9.2** Average loss per compressor (Illustration 9.1)

| *Brand* | *Mean* ($\bar{Y}$) | *Variance* ($\sigma^2$) | *MSD* | $L(Y)$ |
|---|---|---|---|---|
| *A* | 232.2 | 197.56 | 202.4 | 1518.00 |
| *B* | 239.0 | 59.0 | 140.0 | 1050.00 |

Since the average loss per compressor is less for Brand *B*, it is recommended to purchase Brand *B* compressor.

## ILLUSTRATION 9.2

**Smaller–the Better Case**

The problem is concerned with the manufacturing of speedometer cable casings using the raw material supplied by two different suppliers. The output quality characteristic is the shrinkage of the cable casing.

The specified shrinkage ($\Delta_0$) = 2%
Estimated consumer loss ($A_0$) = ₹ 100
The loss function, $L(Y) = KY^2$

$$K = \frac{A_0}{\Delta_0^2} \quad \text{[from Eq. (9.10)]}$$

$$= \frac{100}{(2)^2} = 25$$

The data collected on 20 samples from the two suppliers is given in Table 9.3.

**TABLE 9.3** Data for Illustration 9.2 (Percentage shrinkage)

| | |
|---|---|
| Supplier 1 | 0.20, 0.28, 0.12, 0.20, 0.24, 0.22, 0.32, 0.26, 0.32, 0.20, 0.20, 0.24, 0.20, 0.28, 0.12, 0.20, 0.32, 0.32, 0.22, 0.26 |
| Supplier 2 | 0.15, 0.16, 0.15, 0.10, 0.18, 0.19, 0.11, 0.19, 0.17, 0.18, 0.10, 0.15, 0.16, 0.15, 0.18, 0.19, 0.19, 0.11, 0.18, 0.17 |

The average loss per casing is computed from Eq. (9.11).

$$L(Y) = K(\sigma^2 + \overline{Y}^2)$$

$$= 25(\sigma^2 + \overline{Y}^2)$$

The mean, variance, and the average loss per cable is given in Table 9.4.

**TABLE 9.4** Average loss per cable (Illustration 9.2)

| *Supplier* | *Mean square* ($\overline{Y}^2$) | *Variance* ($\sigma^2$) | $L(Y)$ |
|---|---|---|---|
| 1 | 0.055696 | 0.003424 | 1.478 |
| 2 | 0.04025 | 0.001055 | 0.627 |

The average loss per casing is less with Supplier 2. Hence, Supplier 2 is preferred.

## ILLUSTRATION 9.3

**Larger–the Better Case**

This problem is concerned with the comparison of the life of electric bulbs of same wattage manufactured by two different companies. The quality characteristic ($Y$) is the life of the bulb in hours.

The specified life ($\Delta_0$) = 1000 hr

Consumer loss ($A_0$) = ₹ 10

The data collected is given in Table 9.5.

**TABLE 9.5** Data for Illustration 9.3 (Life of bulbs in hours)

| | |
|---|---|
| Company 1 | 860, 650, 780, 920, 880, 780, 910, 820, 650, 750, 650, 780, 650, 650, 760, 580, 910, 750, 600, 750 |
| Company 2 | 750, 690, 880, 870, 810, 910, 750, 780, 680, 900,750, 900, 940, 690, 670, 800, 810, 820, 750, 700 |

Determine the average loss per bulb for the two companies.

**Solution:**

The loss function, $L(Y) = K(1/Y^2)$

$$K = A_0 \Delta_0^2 \text{ [Eq. (9.13)]}$$
$$= 10(1000)^2$$
$$= 10 \times 10^6$$

The average loss per product is computed from Eq. (9.15) given below.

$$L(Y) = \frac{K}{\overline{Y}^2}\left[1 + \frac{3\sigma^2}{\overline{Y}^2}\right]$$

The computed values are summarized in Table 9.6.

**TABLE 9.6** Average loss per product (Illustration 9.3)

| *Company* | *Mean square* ($\overline{Y}^2$) | *Variance* ($\sigma^2$) | *Average loss per product*, $L(Y)$ |
|---|---|---|---|
| 1 | 5,68,516.00 | 10,894.00 | 18.6 |
| 2 | 6,28,056.25 | 6918.75 | 16.45 |

Since the average loss is less with Company 2, Company 2 is preferred.

## 9.4 ESTIMATION OF QUALITY LOSS

### 9.4.1 Traditional Method

Usually in manufacturing companies, the quality loss is estimated by considering the number of defects/defectives. In this, it is assumed that all parts with in the specification will not have any quality loss.

Loss by defect = Proportion out of specification × number of products × cost per product

### 9.4.2 Quality Loss Function Method

As already discussed, this method is based on dispersion (deviation from target). Depending on the type of quality characteristic, quality loss is estimated using an appropriate quality loss function. Suppose we have a nominal–the best type of quality characteristic.

$$\text{Loss by dispersion} = \frac{A_0}{\Delta_0^2}\,[\sigma^2 + (\overline{Y} - T)^2] \times \text{number of products}$$

**ILLUSTRATION 9.4**

**Comparison of Quality Loss**

A company manufactures a part for which the specification is 40 ± 4 mm (i.e., $T$ = 40 and $\Delta_0$ = 4). The cost of repairing or resetting is ₹ 30 (i.e., $A_0$ = 30). The process average ($\bar{Y}$) is centred at the target $T$ = 40 with a standard deviation of 1.33. The annual production is 50,000 parts.

The proportion out of specification can be estimated from normal distribution as follows. The Upper Specification Limit (USL) is 44 and the Lower Specification Limit (LSL) is 36.

$$Z_{\text{USL}} = \frac{(\text{USL} - \bar{Y})}{\sigma} = \frac{44 - 40}{1.33} = 3.0$$

Therefore, proportion out of specification above USL is 0.00135. Similarly, proportion falling below LSL is 0.00135. Thus, total proportion of product falling out of specification limits is 0.0027 (0.27%).

$$\begin{aligned}\text{Loss by defect} &= \text{Proportion out of specification} \times \text{number of parts} \times \text{cost per part}\\ &= 0.0027 \times 50{,}000 \times 30\\ &= ₹\ 4{,}050.\end{aligned}$$

$$\begin{aligned}\text{Loss by dispersion} &= \frac{A_0}{\Delta_0^2}\,[\sigma^2 + (\bar{Y} - T)^2] \times \text{number of products}\\ &= \frac{30}{4^2}\,[1.33^2 + (40 - 40)^2] \times 50{,}000\\ &= 1.875\ (1.769)\ (50{,}000)\\ &= ₹\ 1{,}65{,}844.\end{aligned}$$

Note the difference between the two estimates. So, to reduce the quality loss we need to reduce the dispersion (variation) there by improving quality.

## PROBLEMS

**9.1** Two processes $A$ and $B$ are used to produce a part. The following data has been obtained from the two processes.

| *Process* | *A* | *B* |
|---|---|---|
| Mean | 100.00 | 105.00 |
| Standard deviation | 13.83 | 10.64 |

The specification for the part is 100 ± 10. The consumer loss was estimated to be ₹ 20. Determine which process is economical based on quality loss.

**9.2** A customer wants to compare the life of electric bulbs of four different brands ($A$, $B$, $C$ and $D$) of same wattage and price. The output characteristic for life ($Y_0$) = 1000 hours and the consumer loss $A_0$ = ₹ 10. The data on the life of bulbs has been obtained for the four brands which are given in Table 9.7. Which brand would you recommend based on quality loss?

**TABLE 9.7** Data on life of bulbs

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $A$ | 800 | 850 | 900 | 750 | 950 | 700 | 800 | 750 | 900 | 800 |
| $B$ | 950 | 750 | 850 | 900 | 700 | 790 | 800 | 790 | 900 | 950 |
| $C$ | 860 | 650 | 780 | 920 | 880 | 780 | 910 | 820 | 650 | 750 |
| $D$ | 750 | 690 | 880 | 870 | 810 | 910 | 750 | 780 | 680 | 900 |

**9.3** Television sets are made with a desired target voltage of 230 volts. If the output voltage lies outside the range of 230 ± 20 V, the customer incurs an average cost of ₹ 200.00 per set towards repair/service. Find the cost coefficient $K$ and establish the quality loss function.

CHAPTER 10

# Taguchi Methods

## 10.1 INTRODUCTION

We know that a full factorial design is one in which all possible combinations of the various factors at different levels are studied. While factorial experimentation is very useful and is based on strong statistical foundations, it becomes unmanageable in the usual industrial context. With increase in the number of factors and their levels, the number of experiments would be prohibitively large. And for conducting so many experiments a number of batches of materials, different process conditions, etc. results in heterogeneity and the experimental results tend to become inaccurate (results in more experimental error). Hence care must be taken that variations in the experimental material, background conditions, etc. do not bias the conclusions to be drawn. To address these issues, statisticians have developed Fractional Replicate Designs (Fractional Factorial Designs) which were discussed in Chapter 7. However, construction of fractional replicate designs generally requires good statistical knowledge on the part of the experimenter and is subject to some constraints that limit the applicability and ease of conducting experiments in practice.

## 10.2 TAGUCHI METHODS

Dr. Genich Taguchi, a Japanese scientist contributed significantly to the field of quality engineering. His quality philosophy is that quality should be designed into the product and not inspected into it. That is, quality is not achieved through inspection which is a postmortem activity. His second philosophy is that quality can be achieved by minimizing the deviation from the target value. And that product design should be such that its performance is insensitive to uncontrollable (noise) factors. He advocated that the cost of quality should be measured as function of deviation from the standard (target). Taguchi Methods/Techniques are

- Off-line Quality Assurance techniques
- Ensures Quality of Design of Processes and Products
- Robust Design is the procedure used
- Makes use of Orthogonal Arrays for designing experiments

### 10.2.1 Development of Orthogonal Designs

Dr. Genich Taguchi suggested the use of Orthogonal Arrays (OAs) for designing the experiments. He has also developed the concept of linear graph which simplifies the design of OA experiments. These designs can be applied by engineers/scientists without acquiring advanced statistical knowledge. The main advantage of these designs lies in their simplicity, easy adaptability to more complex experiments involving number of factors with different numbers of levels. They provide the desired information with the least possible number of trials and yet yield reproducible results with adequate precision. These methods are usually employed to study main effects and applied in screening/pilot experiments.

The resource difference in terms number of experiments conducted between OA experiments and full factorial experiments is given in Table 10.1.

**TABLE 10.1** Comparison of number of experiments in full factorial and OA designs

| *Number of factors* | *Number of levels* | *Number of experiments* | |
|---|---|---|---|
| | | *Full factorial* | *Taguchi* |
| 3 | 2 | 8 | 4 |
| 7 | 2 | 128 | 8 |
| 15 | 2 | 32,768 | 16 |
| 4 | 3 | 81 | 9 |
| 13 | 3 | 1,594,323 | 27 |

Note that the number of experiments conducted using Taguchi methods is very minimal.

## 10.3 ROBUST DESIGN

Robust design process consists of three steps, namely, System Design, Parameter Design and Tolerance Design.

### 10.3.1 System Design

System design is the first step in the design of any product. This involves both the conceptual and functional design of the product. The conceptual design is the creation, exploration and presentation of ideas. It is more about how a product should look like and perform. At the same time the conceived conceptual design must satisfy both the manufacturer and the customer. The functional design involves the identification of various sub tasks and their integration to achieve the functional performance of the end product. In functional design, usually a prototype design (physical or mathematical) is developed by applying scientific and engineering knowledge. Functional design is highly a creative step in which the designers creative ideas and his experience play a vital role. The designer should also keep in mind the weight and cost of the product versus the functional performance of the product.

### 10.3.2 Parameter Design

Parameter design is the process of investigation leading to the establishment of optimal settings of the design parameters so that the product/process perform on target and is not influenced by the noise factors. Statistically designed experiments and/or Orthogonal experiments are used for this purpose.

### 10.3.3 Tolerance Design

Tolerance design is the process of determining the tolerances around the nominal settings identified in parameter design process. Tolerances should be set such that the performance of the product is on target and at the same time they are achievable at minimum manufacturing cost. The optimal tolerances should be developed in order to minimize the total costs of manufacturing and quality. Taguchi suggested the application of quality loss function to arrive at the optimal tolerances.

## 10.4 BASIS OF TAGUCHI METHODS

The basis of Taguchi methods is the Additive Cause-Effect model. Suppose we have two factors ($A$ and $B$) which influence a process. Let $\alpha$ and $\beta$ are the effects of the factors $A$ and $B$ respectively on the response variable $Y$. Taguchi pointed out that in many practical situations these effects (main effects) can be represented by an additive cause-effect model. The additive model has the form

$$Y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \tag{10.1}$$

where,

$\mu$ = mean value of $Y$ in the region of experiment
$\alpha_i$ and $\beta_j$ = individual or main effects of the influencing factors $A$ and $B$
$e_{ij}$ = error term.

The term main effect designates the effect on the response $Y$ that can be attributed to a single process or design parameter, such as $A$. In the additive model, it is assumed that interaction effects are absent. The additivity assumption also implies that the individual effects are separable. Under this assumption, the effect of each factor can be linear, quadratic or higher order. Interactions make the effects of the individual factors non-additive. When interaction is included, the model will become multiplicative (non-additive). Sometimes one is able to convert the multiplicative (some other non-additive model) into an additive model by transforming the response $Y$ into log ($Y$) or $1/Y$ or $\sqrt{Y}$ .

In Taguchi Robust design procedure we confine to the main effects model (additive) only. The experimenter may not know in advance whether the additivity of main effects holds well in a given investigation. One practical approach recommended by Taguchi is to run a confirmation experiment with factors set at their optimal levels. If the result obtained (observed average response) from the confirmation experiment is comparable to the predicted response based on the main effects model, we can conclude that the additive model holds good. Then this model can be used to predict the response for any combination of the levels of the factors, of course, in the region of experimentation. In Taguchi Methods we generally use the term optimal levels which in real sense they are not optimal, but the best levels.

## 10.5 STEPS IN EXPERIMENTATION

The following are steps related to the Taguchi-based experiments:

1. State the problem
2. Determine the objective
3. Determine the response and its measurement
4. Identify factors influencing the performance characteristic
5. Separate the factors into control and noise factors
6. Determine the number of levels and their values for all factors
7. Identify control factors that may interact
8. Select the Orthogonal Array
9. Assign factors and interactions to the columns of OA
10. Conduct the experiment
11. Analyse the data
12. Interpret the results
13. Select the optimal levels of the significant factors
14. Predict the expected results
15. Run a confirmation experiment

Most of these steps are already discussed in Chapter 2. Other steps related to the Taguchi-based experiments are discussed in Chapters 11 and 12.

## PROBLEMS

**10.1** What is the need for Taguchi Methods?

**10.2** What are the advantages of Taguchi Methods over the traditional experimental designs?

**10.3** Explain the three steps of robust design process.

**10.4** State the additional experimental steps involved in robust design experimentation.

CHAPTER

# 11 Design of Experiments using Orthogonal Arrays

## 11.1 INTRODUCTION

Orthogonal Arrays (OAs) were mathematical invention recorded in early 1897 by Jacques Hadamard, a French mathematician. The Hadamard matrices are identical mathematically to the Taguchi matrices; the columns and rows are rearranged (Ross 2005).

### *Nomenclature of arrays*

$L_a(b^c)$
- $L$ = Latin square
- $a$ = number of rows
- $b$ = number of levels
- $c$ = number of columns (factors)

Degrees of freedom associated with the OA = $a - 1$
Some of the standard orthogonal arrays are listed in Table 11.1.

**TABLE 11.1** Standard orthogonal arrays

| *Two-level series* | *Three-level series* | *Four-level series* | *Mixed-level series* |
|---|---|---|---|
| $L_4$ $(2^3)$ | $L_9$ $(3^4)$ | $L_{15}$ $(4^5)$ | $L_{18}$ $(2^1, 3^7)^\dagger$ |
| $L_8$ $(2^7)$ | $L_{27}$ $(3^{13})$ | $L_{64}$ $(4^{21})$ | $L_{36}$ $(2^{11}, 3^{12})$ |
| $L_{16}$ $(2^{15})$ | $L_{81}$ $(3^{40})$ | | |
| $L_{32}$ $(2^{31})$ | | | |
| $L_{12}$ $(2^{11})$* | | | |

* Interactions cannot be studied
† Can study one interaction between the 2-level factor and one 3-level factor

An example of $L_8(2^7)$ OA is given in Table 11.2.

**TABLE 11.2** Standard $L_8$ orthogonal array

| *Trial no.* | *Columns* 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 |

- The 1s and 2s in the matrix indicate the low and high level of a factor respectively.
- There are eight experimental runs (trials) 1 to 8.
- Each column has equal number of 1s and 2s.
- Any pair of columns have only four combinations (1, 1), (1, 2), (2, 1), and (2, 2) indicating that the pair of columns are orthogonal.
- This OA can be used to study upto seven factors.

## 11.2 ASSIGNMENT OF FACTORS AND INTERACTIONS

When the problem is to study only the main factors, the factors can be assigned in any order to each column of the OA. When we have main factors and some interactions to be studied, we have to follow certain procedure. Taguchi has provided the following two tools to facilitate the assignment of factors and interactions to the columns of Orthogonal Arrays.

1. Interaction Tables (Appendix B)
2. Linear Graphs (Appendix C)

The interaction table contains all possible interactions between columns (factors). One interaction table for the two-level series and another one for three-level series are given in Appendix B. Table 11.3 gives part of the two-level interaction table that can be used with the $L_8$ OA. Suppose factor $A$ is assigned to Column 2 and factor $B$ is assigned to Column 4, the interaction $AB$ should be assigned to Column 6 as given in Table 11.3.

**TABLE 11.3** Interaction table for $L_8$ OA

| *Column no.* | *Column no.* 2 | 3 | 4($B$) | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| 1 | 3 | 2 | 5 | 4 | 7 | 6 |
| 2($A$) | | 1 | 6($AB$) | 7 | 4 | 5 |
| 3 | | | 7 | 6 | 5 | 4 |
| 4 | | | | 1 | 2 | 3 |
| 5 | | | | | 3 | 2 |
| 6 | | | | | | 1 |

### 11.2.1 Linear Graph

Taguchi suggested the use of additive effects model in robust design experiments. In the additive model, we include only main effects and do not consider interaction effects. And we conduct a confirmation experiment to determine whether the main factors model is adequate.

Sometimes a few two factor interactions may be important. Taguchi suggested their inclusion in the orthogonal experiments. To simplify the assignment of these interactions to the columns of the OA, Taguchi developed linear graphs.

The Linear Graph (LG) is a graphic representation of interaction information. These are prepared using part of the information from the interaction table. These linear graphs facilitate the assignment of main factors and interactions to the different columns of an OA. Figure 11.1 shows one of the two standard linear graphs associated with $L_8$ OA.

In Figure 11.1, the numbers indicate the column numbers of $L_8$ OA. Suppose we assign main factors *A*, *B*, *C* and *D* to columns 1, 2, 4 and 7 respectively. Then the interactions *AB*, *AC*, and *BC* should be assigned to columns 3, 5 and 6 respectively.

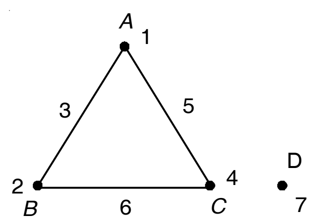



**FIGURE 11.1** Standard linear graph of $L_8$ OA.

## 11.3 SELECTION AND APPLICATION OF ORTHOGONAL ARRAYS

The following steps can be followed for designing an OA experiment (Matrix Experiment):

*Step* 1: Determine the df required for the problem under study.
*Step* 2: Note the levels of each factor and decide the type of OA. (two-level or three-level)
*Step* 3: Select the particular OA which satisfies the following conditions.
(a) degrees of freedom of OA > df required for the experiment
Note that the degrees of freedom of OA = number of rows in the OA minus one.
(b) Possible number of interactions of OA > the number of interactions to be studied.
*Step* 4: Draw the required linear graph for the problem.
*Step* 5: Compare with the standard linear graph of the chosen OA.
*Step* 6: Superimpose the required LG on the standard LG to find the location of factor columns and interaction columns.
The remaining columns (if any) are left out as vacant.
*Step* 7: Draw the layout indicating the assignment of factors and interactions.

The rows will indicate the number of experiments (trials) to be conducted.

**ILLUSTRATION 11.1**

Consider an experiment with four factors (*A*, *B*, *C* and *D*) each at two levels. Also, the interactions *AB*, *AD* and *BD* are of interest to the experimenter. The design of this OA experiment is explained following the procedure given in Section 11.3.

*Step* 1: The required degree of freedom (Table 11.4).

**TABLE 11.4** Degrees of freedom for Illustration 11.1

| *Factors* | *Levels* | *Degrees of freedom* |
|---|---|---|
| $A$ | 2 | $2 - 1 = 1$ |
| $B$ | 2 | $2 - 1 = 1$ |
| $C$ | 2 | $2 - 1 = 1$ |
| $D$ | 2 | $2 - 1 = 1$ |
| $AB$ | | $(2 - 1)(2 - 1) = 1$ |
| $AD$ | | $(2 - 1)(2 - 1) = 1$ |
| $BD$ | | $(2 - 1)(2 - 1) = 1$ |
| | Total | $df = 7$ |

*Step* 2: Levels of factors
All factors are to be studied at two-levels. Hence, choose a two-level OA

*Step* 3: Selection of required OA
(a) The OA which satisfies the required degrees of freedom is $L_8$ OA.
(b) Number of interactions to be studied = 3
Interactions possible in $L_8$ = 3 [Figure 11.3(a)]
Therefore, the best OA would be $L_8$.

*Step* 4: Required linear graph
Figure 11.2 shows the required linear graph for Illustration 11.1.

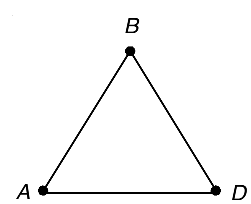



**FIGURE 11.2** Required linear graph for Illustration 11.1.

*Step* 5: The two types of standard linear graphs of $L_8$ OA are shown in Figures 11.3(a) and 11.3(b).

*Step* 6: Superimpose the required linear graph on the standard linear graph.

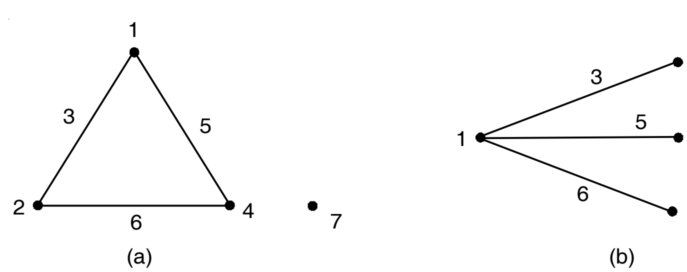



**FIGURE 11.3** Types of standard linear graph of $L_8$ OA.

The standard linear graph shown in Figure 11.3(a) is similar to the required linear graph. The superimposed linear graph for Illustration 11.1 is shown in Figure 11.4.

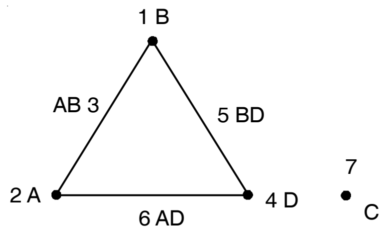



**FIGURE 11.4** Superimposed linear graph for Illustration 11.1.

*Step* 7: The assignment of the factors and interactions to the columns of OA as per the linear graph (Figure 11.4) is shown in the design layout (Table 11.5).

**TABLE 11.5** The design layout for Illustration 11.1

| *Trial no.* | *Factors* | | | | | | | *Response Y* |
|---|---|---|---|---|---|---|---|---|
| | *B* 1 | *A* 2 | *AB* 3 | *D* 4 | *BD* 5 | *AD* 6 | *C* 7 | |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | * * |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | * * |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | * * |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | * * |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | * * |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | * * |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | * * |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | * * |

### *Test sheet*

For conducting the experiment test sheet may be prepared without the interacting columns (Table 11.6). During experimentation, the levels are set only to the main factors and hence interacting columns can be omitted in the test sheet. However, interacting columns are required for data analysis.

Since all the columns of OA is assigned with the factors and interactions, the design is called a saturated design. $L_4$ can be used to study up to three factors. If we want to study more than three factors but up to seven factors, we can use $L_8$ OA. Suppose we have only five factors/interactions to be studied. For this also we can use $L_8$ OA. However, two of the seven columns remain unassigned. These vacant columns are used to estimate error.

**TABLE 11.6** Test sheet for Illustration 11.1 experiment

| *Trial no.* | *Factors* | | | | *Response (Y)* | |
|---|---|---|---|---|---|---|
| | *B* | *A* | *D* | *C* | $R_1$ | $R_2$ |
| | 1 | 2 | 4 | 7 | | |
| 1 | 1 | 1 | 1 | 1 | * | * |
| 2 | 1 | 1 | 2 | 2 | * | * |
| 3 | 1 | 2 | 1 | 2 | * | * |
| 4 | 1 | 2 | 2 | 1 | * | * |
| 5 | 2 | 1 | 1 | 2 | * | * |
| 6 | 2 | 1 | 2 | 1 | * | * |
| 7 | 2 | 2 | 1 | 1 | * | * |
| 8 | 2 | 2 | 2 | 2 | * | * |

*R*–Replication

## ILLUSTRATION 11.2

An arc welding process is to be studied to determine the best levels for the process parameters in order to maximize the mechanical strength of the welded joint. The factors and their levels for study have been identified and are given in Table 11.7.

**TABLE 11.7** Factors and their levels for Illustration 11.2

| *Factors* | *Levels* | |
|---|---|---|
| | 1 | 2 |
| Type of welding rod (A) | B10 | J100 |
| Weld material (B) | SS41 | SB35 |
| Thickness of material (C) | 8 mm | 12 mm |
| Angle of welded part (D) | 70° | 75° |
| Current (E) | 130 A | 150 A |
| Preheating (F) | No | Yes |
| Welding method (G) | Single | Weaving |
| Cleaning method (H) | Wire brush | Grinding |

In addition to the main effects, the experimenter wanted to study the two-factor interactions, *AC*, *AH*, *AG*, and *GH*. The design of OA experiment for studying this problem is explained below.

***Design of the experiment:*** All factors are to be studied at two-levels. So, a two-level series OA shall be selected.

Degrees of freedom required for the problem are

Main factors = 8 (each factor has one degree of freedom)
Interactions = 4
Total df = 12

To accommodate 12 df, we have to use $L_{16}(2^{15})$ OA.

The required linear graph for Illustration 11.2 is shown in Figure 11.5.

The required linear graph is closer to part of the standard linear graph as shown in Figure 11.6.

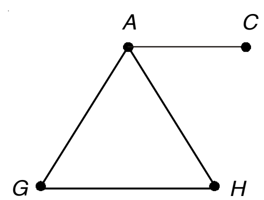



**FIGURE 11.5** Required linear graph for Illustration 11.2.

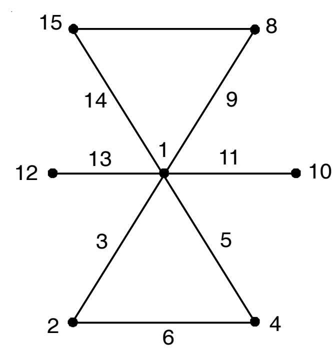



**FIGURE 11.6** Standard linear graph of $L_{16}$ OA.

From the standard linear graph (Figure 11.6), the part which is similar to the required linear graph is selected and superimposed (Figure 11.7). The assignment of factors and interactions to the columns of $L_{16}(2^{15})$ OA is given in Table 11.8. The three unassigned columns marked with $e$ are used to estimate the error.

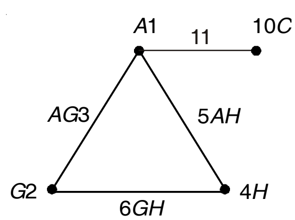



**FIGURE 11.7** Superimposed linear graph for Illustration 11.2.

Linear graphs facilitate easy and quick assignment of factors and interactions to the columns of Orthogonal Arrays. Sometimes, the required linear graph may not match with the standard linear graph or may partly match. In such cases, along with the partly matched linear graph, we use the triangular table (Appendix B) to complete the assignment. Standard linear graphs are not given for all the Orthogonal Arrays. When standard linear graph is not available, the triangular table is used.

Additional information on Orthogonal Arrays and linear graphs is available in Phadke (2008).

**TABLE 11.8** Assignment of factors and interactions for Illustration 11.2

| *Trial no.* | 1 *A* | 2 *G* | 3 *AG* | 4 *B* | 5 *D* | 6 *E* | 7 *F* | 8 *H* | 9 *AH* | 10 *GH* | 11 *e* | 12 *C* | 13 *AC* | 14 *e* | 15 *e* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 |
| 4 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 |
| 5 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 |
| 6 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 |
| 7 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 |
| 8 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 |
| 9 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| 10 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 |
| 11 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 |
| 12 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 |
| 13 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 |
| 14 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 |
| 15 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 |
| 16 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 |

## PROBLEMS

**11.1** An engineer wants to study the effect of the control factors *A*, *B*, *C*, *D* and *E* including the interactions *AB* and *AC* affecting the hardness of a metal component. The objective is to maximize the hardness. Design an OA experiment.

**11.2** An industrial engineer wants to study a manual welding process with factors each at two levels. The factors are welding current (*A*), welding time (*B*), thickness of plates (*C*), type of welding (*D*), and operator (*E*). In addition, he wants to study interactions *AB*, *AE*, *BD*, *BE*, *CE* and *DE*. Design an OA experiment.

**11.3** An experimenter wants to study the effect of five main factors *A*, *B*, *C*, *D*, and *E* each at two-level and two-factor interactions *AC*, *BC*, *AD*, *AE*, and *BE*. Design an OA experiment.

**11.4** An experimenter wants to study the effect of five main factors *A*, *B*, *C*, *D* and *E* each at two-level and two-factor interactions *AB*, *AD*, *BC*, *BD*, *ED* and *CE*. Design an OA experiment.

**11.5** An engineer performed an experiment on the control factors *A*, *B*, *C*, *D* and *E*, including the interactions *AB*, *BC*, *BD*, *CD* and *AC*, affecting the hardness of a metal component. The objective is to maximize the hardness. Design an OA experiment.

**11.6** A heat treatment process was suspected to be the cause for higher rejection of gear wheels. Hence, it was decided to study the process by Taguchi method. The factors considered for the experiment and their levels are given in Table 11.9.

**TABLE 11.9** Factors and their levels for Problem 11.6

| *Factors* | *Levels* | |
|---|---|---|
| | 1 | 2 |
| Hardening temperature (*A*) | 830°C | 860°C |
| First quenching time (*B*) | 50 s | 60 s |
| Second quenching time (*C*) | 15 s | 20 s |
| Quenching media (*D*) | Fresh water | Salt water |
| Quenching method (*E*) | Top up | Top down |
| Preheating (*F*) | No | Yes, 450°C |
| Tempering temperature (*G*) | 190°C | 220°C |

In addition, the interaction effects *AB*, *AC*, *AF*, *BD*, *BE*, *CD* and *CE* are required to be studied. Design a Taguchi experiment for conducting this study.

CHAPTER

# 12 Data Analysis from Taguchi Experiments

## 12.1 INTRODUCTION

Data collected from Taguchi/Orthogonal Array (OA) experiments can be analysed using response graph method or Analysis of Variance (ANOVA).

The response graph method is very easy to understand and apply. This method requires no statistical knowledge. For practical/industrial applications, this method may be sufficient.

Analysis of variance (ANOVA) has already been discussed in the earlier chapters. This method accounts the variation from all sources including error term. If error sum of squares is large compared to the control factors in the experiment, ANOVA together with percent contribution indicate that the selection of optimum condition may not be useful. Also for statistically validating the results, ANOVA is required. In this chapter how various types of data are analysed is discussed.

## 12.2 VARIABLE DATA WITH MAIN FACTORS ONLY

The design of OA experiments has been discussed in Chapter 11. By conducting the experiment we obtain the response (output) from each experiment. The response obtained from all the trials of an experiment is termed as data. The data can be either attribute or variable. We know that variable data is obtained by measuring the response with an appropriate measurement system. Suppose in an OA experiment we have only the main factors and the response is a variable. The analysis of data from these types of experiments is discussed in the following illustration.

**ILLUSTRATION 12.1**

An experiment was conducted to study the effect of seven main factors (*A*, *B*, *C*, *D*, *E*, *F*, and *G*) each at two levels using $L_8$ OA. The results (response) from two replications are given in Table 12.1. The data analysis from this experiment is discussed as follows:

### *Data analysis using response graph method*

The following steps discuss data analysis using response graph method:

*Step* 1: Develop response totals table for factor effects (Table 12.2).

**TABLE 12.1** Data for Illustration 12.1

| *Trial no.* | *Factors/Columns* | | | | | | | *Results* | |
|---|---|---|---|---|---|---|---|---|---|
| | *A* 1 | *B* 2 | *C* 3 | *D* 4 | *E* 5 | *F* 6 | *G* 7 | $R_1$ | $R_2$ |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 11 | 11 |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 4 | 4 |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 4 | 10 |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 4 | 8 |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 9 | 4 |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 4 | 3 |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 4 |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 10 | 8 |

**TABLE 12.2** Response totals for Illustration 12.1

| Factors | *A* | *B* | *C* | *D* | *E* | *F* | *G* |
|---|---|---|---|---|---|---|---|
| Level 1 | 56 | 50 | 53 | 54 | 61 | 65 | 46 |
| Level 2 | 43 | 49 | 46 | 45 | 38 | 34 | 53 |

This table is developed by adding the response values corresponding to each level (Level 1 and Level 2) of each factor. For example, Level 1 totals of factor *A* is sum of the observations from trials (runs) 1, 2, 3 and 4 (Table 12.1). That is,

$$A_1 = (11 + 11) + (4 + 4) + (4 + 10) + (4 + 8) = 56$$

Similarly, the Level 2 totals of factor *B* is the sum of response values from trials 3, 4, 7 and 8. That is,

$$B_2 = (4 + 10) + (4 + 8) + (1 + 4) + (10 + 8) = 49$$

Thus there are 8 observations in each total.

*Step* 2: Construct average response table and rank the level difference of each factor.

The response totals are converted into average response and given in Table 12.3. Each total (Table 12.2) is divided by 8 to calculate average. The absolute difference in the average response of the two levels of each factor is also recorded. This difference represents the effect of the factor. These differences are ranked starting with the highest difference as rank 1, the next highest difference as rank 2 and so on. Ties, if any are arbitrarily broken.

**TABLE 12.3** Average response and ranking of factor effects

| *Factors* | *A* | *B* | *C* | *D* | *E* | *F* | *G* |
|---|---|---|---|---|---|---|---|
| Level 1 | 7.00 | 6.25 | 6.63 | 6.75 | 7.63 | 8.13 | 5.75 |
| Level 2 | 5.38 | 6.13 | 5.75 | 5.63 | 4.75 | 4.25 | 6.63 |
| Difference | 1.62 | 0.12 | 0.88 | 1.12 | 2.88 | 3.88 | 0.88 |
| Rank | 3 | 7 | 5 | 4 | 2 | 1 | 6 |

From Table 12.3, it is observed that factor $F$ has the largest effect (rank 1). And the grand average/overall mean,

$$\bar{Y} = \frac{\text{Grand total of all observations}}{\text{Total number of observations}} = \frac{99}{16} = 6.19$$

*Step* 3: Draw the response graph.

The response graph with the average values is shown in Figure 12.1.

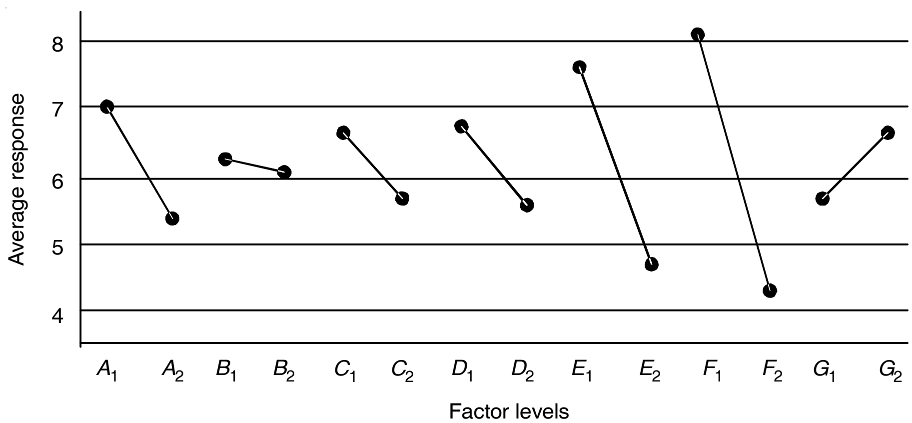



**FIGURE 12.1** Response graph for Illustration 12.1.

*Step* 4: Predict the optimum condition.

Based on the objective of the experiment, that is whether minimization of the response or maximization, the optimum condition is selected. Suppose in this experiment the objective is maximization of response. For maximization of response, the optimum condition is selected based on higher mean value of each factor. Accordingly from Figure 12.1, the optimum condition is given by

$$A_1\, B_1\, C_1\, D_1\, E_1\, F_1\, G_2$$

Generally, all factors will not contribute significantly. As a rule of thumb, it is suggested to take the number of significant effects (factors) equal to about one-half the number of degrees of freedom of the OA used for the experiment. These are selected in the order of their ranking starting from rank 1. In this example, we consider 3 effects (rank 1, rank 2 and rank 3) as significant. Thus, the optimum condition is $F_1\, E_1\, A_1$.

The predicted optimum response $\mu_{\text{pred}}$ is given by

$$\mu_{\text{pred}} = \bar{Y} + (\bar{F}_1 - \bar{Y}) + (\bar{E}_1 - \bar{Y}) + (\bar{A}_1 - \bar{Y})$$

$$= (\bar{F}_1 + \bar{E}_1 + \bar{A}_1) - 2\bar{Y} \qquad (12.1)$$

where,

$\bar{Y}$ = overall mean response and

$\bar{F}_1, \bar{E}_1, \bar{A}_1$ = average response at Level 1 of these factors.

$$\begin{aligned}\mu_{\text{pred}} &= 6.19 + (8.13 - 6.19) + (7.63 - 6.19) + (7.00 - 6.19)\\ &= (8.13 + 7.63 + 7.00) - 2(6.19)\\ &= 22.76 - 12.38 = 10.38\end{aligned}$$

### *Data analysis using analysis of variance*

First the correction factor (CF) is computed.

$$\text{CF} = \frac{T^2}{N}$$

where,

$T$ = grand total and
$N$ = total number of observations

$$\text{CF} = \frac{(99)^2}{16} = 612.56$$

$SS_{\text{Total}}$ is computed using the individual observations (response) data as already discussed.

$$\begin{aligned}SS_{\text{Total}} &= \sum_{i=1}^{N} Y_i^2 - \text{CF} \qquad (12.2)\\ &= (11)^2 + (11)^2 + (4)^2 + \ldots + (10)^2 + (8)^2 - 612.56\\ &= 160.44\end{aligned}$$

The factor (effect) sum of squares is computed using the level totals (Table 12.1).

$$SS_A = \frac{A_1^2}{n_{A1}} + \frac{A_2^2}{n_{A2}} - \text{CF} \qquad (12.3)$$

where,

$A_1$ = level 1 total of factor $A$
$n_{A1}$ = number of observations used in $A_1$
$A_2$ = level 2 total of factor $A$
$n_{A2}$ = number of observations used in $A_2$

$$SS_A = \left(\frac{(56)^2}{8} + \frac{(43)^2}{8}\right) - 612.56 = 10.57$$

Similarly, the sum of squares of all effects are computed

$$SS_B = \left(\frac{(50)^2}{8} + \frac{(49)^2}{8}\right) - 612.56 = 0.07$$

$$SS_C = \left(\frac{(53)^2}{8} + \frac{(46)^2}{8}\right) - 612.56 = 3.07$$

$$SS_D = \left(\frac{(54)^2}{8} + \frac{(45)^2}{8}\right) - 612.56 = 5.07$$

$$SS_E = \left(\frac{(61)^2}{8} + \frac{(38)^2}{8}\right) - 612.56 = 33.07$$

$$SS_F = \left(\frac{(65)^2}{8} + \frac{(34)^2}{8}\right) - 612.56 = 60.07$$

$$SS_G = \left(\frac{(46)^2}{8} + \frac{(53)^2}{8}\right) - 612.56 = 3.07$$

The error sum of squares is calculated by subtracting the sum of all factor sums of squares from the total sum of squares.

$$SS_e = SS_{\text{Total}} - (SS_A + SS_B + SS_C + SS_D + SS_E + SS_F + SS_G) \tag{12.4}$$
$$= 160.44 - (10.57 + 0.07 + 3.07 + 5.07 + 33.07 + 60.07 + 3.07)$$
$$= 45.45$$

This error sum of squares is due to replication of the experiment, is called *experimental error* or *pure error*. The initial ANOVA with all the effects is given in Table 12.4. It is to be noted that when we compute the sum of squares of any assigned factor/interaction column (for example, $SS_A$), we are, in fact, computing the sum of squares of that particular column. That is, $SS_A$ = SS of Column 1. Also, the total sum of squares is equal to the sum of squares of all columns.

$$SS_{\text{Total}} = SS_{\text{columns}}$$

Thus, if we have some unassigned columns,

$$SS_{\text{Total}} = SS_{\text{assigned columns}} + SS_{\text{unassigned columns}} \tag{12.5}$$

**TABLE 12.4** ANOVA (initial) for Illustration 12.1

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ | *C*(%) | *Rank* |
|---|---|---|---|---|---|---|
| *A* | 10.57 | 1 | 10.57 | 1.86 | 6.59 | 3 |
| *B* | 0.07 | 1 | 0.07 | 0.01 | 0.04 | 7 |
| *C* | 3.07 | 1 | 3.07 | 0.54 | 1.91 | 5 |
| *D* | 5.07 | 1 | 5.07 | 0.89 | 3.16 | 4 |
| *E* | 33.07 | 1 | 33.07 | 5.82 | 20.62 | 2 |
| *F* | 60.07 | 1 | 60.07 | 10.58 | 37.44 | 1 |
| *G* | 3.07 | 1 | 3.07 | 0.54 | 1.91 | 6 |
| Error (pure) | 45.45 | 8 | 5.68 | 28.33 | | |
| Total | 160.44 | 15 | | | | |

When experiment is not replicated, the sum of squares of unassigned columns is treated as error sum of squares. Even when all columns are assigned and experiment not replicated, we can obtain error variance by pooling the sum of squares of small factor/interaction variances. When all columns are not assigned and experiment is replicated, we will have both experimental error (due to replication) and error from unassigned columns. These two errors can be combined together to get more degrees of freedom and $F$-tested the effects.

***Discussion:*** The $F$-value from table at 5% significance level is $F_{0.05,1,8} = 5.32$. So, from ANOVA (Table 12.4) we see that only factors $E$ and $F$ are significant. When error sum of squares are high and/or error degrees of freedom are less, most of the factors show significance. This type of result may not be useful in engineering sense. To increase the degrees of freedom for error sum of squares, we pool some of the factor/interaction sum of squares so that in the final ANOVA, we will have effects equal to one-half of the degrees of freedom of the OA used in the experiment. This is similar to the selection of the effects (based on rank) for determining the optimum condition in the response graph method.

***Pooling of sum of squares:*** There are two approaches suggested, i.e., pooling down and pooling up, for pooling the sum of squares of factors/interactions into the error term.

***Pooling down:*** In the pooling down approach, we test the largest factor variance with the pooled variance of all the remaining factors. If that factor is significant, the next largest factor is removed from the pool and the $F$-test is done on those two factors with the remaining pooled variance. This is repeated until some insignificant $F$ value is obtained.

***Pooling up:*** In the pooling up strategy, the smallest factor variance is $F$-tested using the next larger factor variance. If no significant $F$-exists, these two are pooled together to test the next larger factor effect until some significant $F$ is obtained. As a rule of thumb, pooling up to one-half of the degrees of freedom has been suggested. In the case of saturated design this is equivalent to have one-half of the effects in the ANOVA after pooling.

It has been recommended to use pooling up strategy. Under pooling up strategy, we can start pooling with the lowest factor/interaction variance and the next lowest and so on until the effects are equal to one-half of the degrees of freedom used in the experiment (3 or 4 in this Illustration 12.1). The corresponding degrees of freedom are also pooled. The final ANOVA is given in Table 12.5 after pooling. Generally, the pooled $SS_e$ should not be more than 50% of $SS_{\text{TOTAL}}$ for half the degrees of freedom of OA used in the experiment.

**TABLE 12.5** ANOVA (final) for Illustration 12.1

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean squares* | $F_0$ | *C(%)* |
|---|---|---|---|---|---|
| $A$ | 10.57 | 1 | 10.57 | 2.23 | 6.59 |
| $E$ | 33.07 | 1 | 33.07 | 6.99 | 20.62 |
| $F$ | 60.07 | 1 | 60.07 | 12.70 | 37.44 |
| Error (pooled) | 56.73 | 12 | 4.73 | | 35.35 |
| Total | 160.44 | 15 | | | 100.00 |

As $F_{0.05,1,12} = 4.75$, factors $E$ and $F$ are significant. Note that factors $E$ and $F$ together contribute about 52% of total variation. The optimal levels for the significant factors are selected based on the mean (average) response. For insignificant factors, levels are selected based on economic criteria even though any level can be used. For maximization of response, the optimal condition is $A_1\, E_1\, F_1$. This result is same as that of response graph method.

***Predicting the optimum response:*** The optimum condition is $A_1\, E_1\, F_1$.

$$\mu_{\text{pred}} = \overline{Y} + (\overline{A}_1 - \overline{Y}) + (\overline{E}_1 - \overline{Y}) + (\overline{F}_1 - \overline{Y}) \tag{12.6}$$

$$= \overline{A}_1 + \overline{E}_1 + \overline{F}_1 - 2\overline{Y}$$

Substituting the average response values from Table 12.3, we get

$$\mu_{\text{pred}} = (7.00 + 7.63 + 8.13) - 2(6.19)$$

$$= 10.38$$

***Per cent contribution (C):*** The per cent contribution indicates the contribution of each factor/interaction to the total variation. By controlling the factors with high contribution, the total variation can be reduced leading to improvement of process/product performance. The per cent contribution due to error before pooling (Table 12.4) indicates the accuracy of the experiment. As a rule of thumb, if the per cent contribution due to error is about 15% or less, we can assume that no important factors have been excluded from experimentation. If it is higher (50% or more), we can say that some factors have not been considered during experimentation conditions were not well controlled or there was a large measurement error.

From Table 12.4, it is found that the rank order based on contribution is same as the rank order obtained by the response graph method. Thus, we can conclude that for industrial application any one of these two methods can be employed.

### *Advantages of ANOVA*

The following are the advantages of ANOVA:

1. Sum of squares of each factor is accounted.
2. If $SS_e$ is large compared to the controlled factors in the experiment, ANOVA together with the percent contribution will suggest that there is little to be gained by selecting optimum conditions. This information is not available from the response table. High error contribution also indicates that some factors have not been included in the study.
3. The confidence interval for the predicted optimum process average can be constructed.
4. The results can be statistically validated.

## 12.3 VARIABLE DATA WITH INTERACTIONS

The design of an OA experiment when we have both main factors and interactions has already been discussed in Chapter 11. If variable data is collected from an experiment involving main factors and interactions, the procedure to be adopted for analyzing the data is explained in the following illustration.

**ILLUSTRATION 12.2**

Suppose we have the data from $L_8$ OA experiment with interactions as given in Table 12.6.

**TABLE 12.6** Experimental data with interactions (Illustration 12.2)

| *Trial no.* | *Factors/Columns* | | | | | | | *Response* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *D* 1 | *C* 2 | *CD* 3 | *A* 4 | *AD* 5 | *B* 6 | *E* 7 | $R_1$ | $R_2$ | $R_3$ |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 11 | 4 | 11 |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 4 | 4 | 4 |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 4 | 1 | 14 |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 4 | 0 | 8 |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 9 | 8 | 4 |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 4 | 1 | 1 |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 4 | 4 |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 14 | 4 | 8 |

***Data analysis using response graph method***

The response totals computed from $L_8$ OA (Table 12.6) is given in Table 12.7.

**TABLE 12.7** Response totals for Illustration 12.2

| *Factors* | *A* | *B* | *C* | *D* | *E* | *AD* | *CD* |
|---|---|---|---|---|---|---|---|
| Level 1 | 75 | 85 | 65 | 69 | 53 | 77 | 73 |
| Level 2 | 56 | 46 | 66 | 62 | 78 | 54 | 58 |

The grand average ($\overline{Y}$) = 131/24 = 5.46.

Note that there are 12 observations in each level total.

Table 12.8 gives the average response and the rank order of the absolute difference in average response between the two levels of each factor.

**TABLE 12.8** Average response and ranking of factor effects for Illustration 12.2

| *Factors* | *A* | *B* | *C* | *D* | *E* | *AD* | *CD* |
|---|---|---|---|---|---|---|---|
| Level 1 | 6.25 | 7.08 | 5.42 | 5.75 | 4.42 | 6.42 | 6.08 |
| Level 2 | 4.67 | 3.83 | 5.50 | 5.17 | 6.50 | 4.50 | 4.83 |
| Difference | 1.58 | 3.25 | 0.08 | 0.58 | 2.08 | 1.92 | 1.25 |
| Rank | 4 | 1 | 7 | 6 | 2 | 3 | 5 |

The response graph is shown in Figure 12.2.

***Optimal levels:*** We select the significant effects equal to one-half of degrees of freedom of the OA used in the experiment. Accordingly, we select 3 or 4 as significant effects. Considering 4 effects as significant, we have *B*, *E*, *AD* and *A* (rank 1, 2, 3 and 4) as significant. Since the interaction effect is significant, to find the optimum condition the interaction should be broken

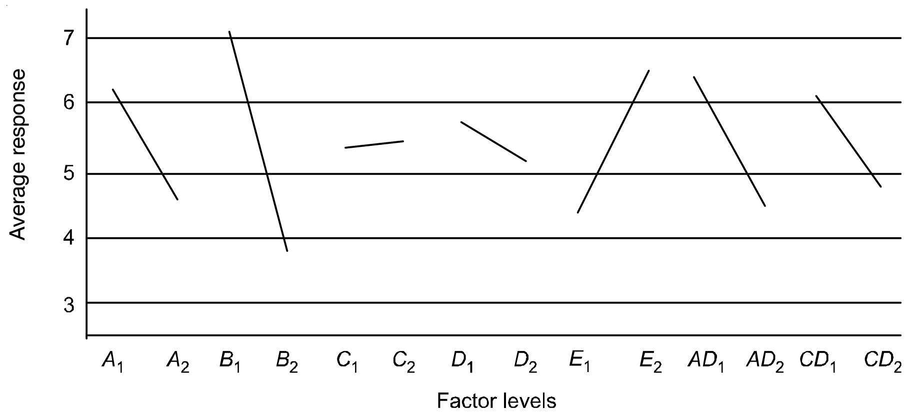



**FIGURE 12.2** Response graph for Illustration 12.2.

down into $A_1 D_1$, $A_1 D_2$, $A_2 D_1$, and $A_2 D_2$; corresponding to the pairs of combinations that occur in a two level OA: (1,1), (1,2), (2,1), and (2,2) respectively. The breakdown totals for the interaction $AD$ is given in Table 12.9. These totals are computed from the response data in $L_8$ OA (Table 12.6). Note that each total is a sum of 6 observations. For maximization of response the optimum interaction component is $A_1 D_1$ (Table 12.9). Hence, for maximization, the optimum condition is

$$B_1,\ E_2,\ A_1 D_1,\ A_1$$

And the optimal levels for the factors is $A_1$, $B_1$, $D_1$ and $E_2$

**TABLE 12.9** *AD* Interaction breakdown response totals

| | $D_1$ | $D_2$ |
|---|---|---|
| $A_1$ | 45 | 24 |
| $A_2$ | 30 | 32 |

The predicted optimum response ($\mu_{\text{pred}}$) is given by

$$\mu_{\text{pred}} = \bar{Y} + (\bar{A}_1 - \bar{Y}) + (\bar{B}_1 - \bar{Y}) + (\bar{E}_2 - \bar{Y}) + [(\overline{A_1D_1} - \bar{Y}) - (\bar{A}_1 - \bar{Y}) - (\bar{D}_1 - \bar{Y})] \quad (12.7)$$

$$= \bar{B}_1 + \bar{E}_2 + \overline{A_1D_1} - \bar{D}_1 - \bar{Y}$$

$$= 7.08 + 6.50 + 7.50 - 5.78 - 5.46 \qquad (\overline{A_1D_1} = 45/6)$$

$$= 9.87$$

***Data analysis using analysis of variance***

The sum of squares is computed using the response totals in Table 12.7.

$$\text{Correction factor (CF)} = \frac{(131)^2}{24} = 715.04$$

$$SS_A = \frac{(75)^2}{12} + \frac{(56)^2}{12} - \text{CF} = 15.04$$

Similarly, the sum of squares of all main effects/factors is obtained.

$$SS_B = 63.38,\ SS_C = 0.04,\ SS_D = 2.04 \text{ and } SS_E = 26.04.$$

$SS_{\text{Total}}$ is computed using the individual responses as usual.

$$SS_{\text{Total}} = (11)^2 + (4)^2 + (11)^2 + (4)^2 + \ldots + (4)^2 + (8)^2 - \text{CF} = 371.97$$

Actually in all OA experiments, when we determine the sum of squares of any effect, it is equivalent to the sum of squares of that column to which that effect is assigned. Therefore, the interaction sum of squares is also computed using the level totals of the interaction column.

$$SS_{AD} = \frac{(77)^2}{12} + \frac{(54)^2}{12} - \text{CF} = 22.04$$

$$SS_{CD} = \frac{(73)^2}{12} + \frac{(58)^2}{12} - \text{CF} = 9.38$$

These computations are summarized in Table 12.10.

**TABLE 12.10** ANOVA (initial) for Illustration 12.2

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ | *C(%)* | *Rank* |
|---|---|---|---|---|---|---|
| *A* | 15.04 | 1 | 15.04 | 1.03 | 4.04 | 4 |
| *B* | 63.38 | 1 | 63.38 | 4.33 | 17.04 | 1 |
| *C* | 0.04 | 1 | 0.04 | 0.00 | 0.01 | 7 |
| *D* | 2.04 | 1 | 2.04 | 0.14 | 0.55 | 6 |
| *E* | 26.04 | 1 | 26.04 | 1.78 | 7.00 | 2 |
| *AD* | 22.04 | 1 | 22.04 | 1.51 | 5.93 | 3 |
| *CD* | 9.38 | 1 | 9.38 | 0.64 | 2.52 | 5 |
| Error (pure) | 234.01 | 16 | 14.63 | | 62.91 | |
| Total | 371.97 | 23 | | | 100.00 | |

***Inference:*** Since $F_{0.05,1,16} = 4.49$, none of the effects are significant. The error variation is very large (contribution = 62.91%), suggesting that some factors would not have been included in the study. However, the rank order based on contribution is same as that obtained earlier (response graph method). Using the pooling rule, we can pool $SS_C$, $SS_D$ and $SS_{CD}$ into the error term leaving four effects in the final ANOVA (Table 12.11) equal to one-half of the number of degrees of freedom of the experiment.

At 5% level of significance, with pooled error variance, factor *B* shows significance. When we are dealing with variable data, we can use either response graph/ranking method or ANOVA. We prefer OA experiments when we are dealing with more number of factors, especially in the early stage of experimentation (screening/pilot experiments). At this stage, usually only main factors are studied. However, we can also study a few important interactions along with the main factors in these designs. As pointed out already, ANOVA is used when we want to validate the results statistically, which will be accepted by all. The optimal levels can be obtained as in the response graph method.

**TABLE 12.11** ANOVA (final) for Illustration 12.2

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ |
|---|---|---|---|---|
| $A$ | 15.04 | 1 | 15.04 | 1.16 |
| $B$ | 63.38 | 1 | 63.38 | 4.90 |
| $E$ | 26.04 | 1 | 26.04 | 2.02 |
| $AD$ | 22.04 | 1 | 22.04 | 1.71 |
| Error (pooled) | 245.47 | 19 | 12.92 | |
| Total | 371.97 | 23 | | |

## 12.4 VARIABLE DATA WITH A SINGLE REPLICATE AND VACANT COLUMN

We use OA designs to have less number of experimental trials leading to savings in resources and time. Some times there may be a constraint on resources prohibiting the replication of the experiment. Also there may be cases where we may not be able to assign with factors and/or interactions to all columns of the OA. With the result, there may be one or more columns left vacant in the OA. The following illustration explains how variable data from an experiment with a single replicate and vacant column(s) is analysed.

**ILLUSTRATION 12.3**

An experiment was conducted using $L_{16}$ OA and data was obtained which is given in Table 12.12.

**TABLE 12.12** Data for Illustration 12.3

| *Trial no.* | 1 $D$ | 2 $BE$ | 3 $F$ | 4 $G$ | 5 $AF$ | 6 $A$ | 7 $CE$ | 8 $E$ | 9 $AC$ | 10 $B$ | 11 $BD$ | 12 $AB$ | 13 $e$ | 14 $CD$ | 15 $C$ | *Response* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.46 |
| 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 0.30 |
| 3 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 0.60 |
| 4 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 0.56 |
| 5 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 0.60 |
| 6 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 0.60 |
| 7 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 0.40 |
| 8 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 0.50 |
| 9 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 0.56 |
| 10 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 0.52 |
| 11 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 0.38 |
| 12 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 0.40 |
| 13 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 0.58 |
| 14 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 0.42 |
| 15 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 0.22 |
| 16 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 0.28 |
| | | | | | | | | | | | | | | | | Total = 7.38 |

Note that the vacant Column 13 is assigned with $e$, denoting error.

***Data analysis:*** Note that we have only one replication and there is one vacant column in Table 12.12. From the vacant column (column 13), we compute error sum of squares. Let us use ANOVA method for Illustration 12.3. The response totals for all the factors and interactions and the vacant column are given in Table 12.13.

**TABLE 12.13** Response totals for the Illustration 12.3

| *Main factors* | *Response total* | | *Interactions and vacant column* | *Response total* | |
|---|---|---|---|---|---|
| | *Level* 1 | *Level* 2 | | *Level* 1 | *Level* 2 |
| $A$ | 3.24 | 4.14 | $AB$ | 3.94 | 3.44 |
| $B$ | 3.8 | 3.58 | $AC$ | 3.68 | 3.70 |
| $C$ | 3.78 | 3.60 | $AF$ | 3.24 | 4.14 |
| $D$ | 4.02 | 3.36 | $BD$ | 3.88 | 3.50 |
| $E$ | 3.80 | 3.58 | $BE$ | 3.78 | 3.60 |
| $F$ | 3.42 | 3.96 | $CD$ | 3.62 | 3.76 |
| $G$ | 4.04 | 3.34 | $CE$ | 3.44 | 3.94 |
| | | | Error ($e_1$) | 3.66 | 3.72 |

***Computation of sum of squares***

Correction factor (CF) = $(7.38)^2/16$ = 3.4040

$$SS_A = \frac{A_1^2}{n_{A1}} + \frac{A_2^2}{n_{A2}} - \text{CF} \tag{12.8}$$

$$= \frac{(3.24)^2}{8} + \frac{(4.14)^2}{8} - \frac{(7.38)^2}{16}$$

$$= 3.4547 - 3.4040 = 0.05063$$

Similarly $SS_B$ = 0.00303, $SS_C$ = 0.00203, $SS_D$ = 0.02723, $SS_E$ = 0.00303, $SS_F$ = 0.01822, $SS_G$ = 0.03063, $SS_{AB}$ = 0.01563, $SS_{AC}$ = 0.00003, $SS_{AF}$ = 0.05063, $SS_{BD}$ = 0.00903, $SS_{BE}$ = 0.00203, $SS_{CD}$ = 0.00122, $SS_{CE}$ = 0.01563 and the vacant column sum of squares $SSe_1$ = 0.00023.

These computations are summarized in Table 12.14.

Now, the effects whose contribution is less ($B$, $C$, $E$, $AC$, $BD$, $BE$ and $CD$) are pooled and added with $SSe_1$ to obtain the pooled error which will be used to test the other effects. This is given in the final ANOVA (Table 12.15).

At 5% significance level, all the effects in Table 12.15 are significant.

***Determination of optimal levels:*** It can be seen from the ANOVA (Table 12.15), the main effects $A$, $D$, $F$ and $G$ and interaction effects $AB$, $AF$ and $CE$ are significant at 5% significance level. The average response for all the main effects is given in Table 12.16. For each significant interaction effect, the average response for the four combinations is given in Table 12.17.

**TABLE 12.14** ANOVA (initial) for the Illustratrion 12.3

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean square* | *C(%)* |
|---|---|---|---|---|
| *A* | 0.05063 | 1 | 0.05063 | 22.09 |
| *B* | 0.00303 | 1 | 0.00303 | 1.32 |
| *C* | 0.00203 | 1 | 0.00203 | 0.88 |
| *D* | 0.02723 | 1 | 0.02723 | 11.88 |
| *E* | 0.00303 | 1 | 0.00303 | 1.32 |
| *F* | 0.01822 | 1 | 0.01822 | 7.95 |
| *G* | 0.03063 | 1 | 0.03063 | 13.36 |
| *AB* | 0.01563 | 1 | 0.01563 | 6.82 |
| *AC* | 0.00003 | 1 | 0.00003 | 0.01 |
| *AF* | 0.05063 | 1 | 0.05063 | 22.09 |
| *BD* | 0.00903 | 1 | 0.00903 | 3.94 |
| *BE* | 0.00203 | 1 | 0.00203 | 0.88 |
| *CD* | 0.00122 | 1 | 0.00122 | 0.54 |
| *CE* | 0.01563 | 1 | 0.01563 | 6.82 |
| Error ($SSe_1$) | 0.00023 | 1 | 0.00023 | 0.10 |
| Total | 0.22923 | 15 | | 100.00 |

**TABLE 12.15** ANOVA (final) for the Illustration 12.3

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ | *C(%)* |
|---|---|---|---|---|---|
| *A* | 0.05063 | 1 | 0.05063 | 19.62 | 22.09 |
| *D* | 0.02723 | 1 | 0.02723 | 10.55 | 11.88 |
| *F* | 0.01822 | 1 | 0.01822 | 7.06 | 7.95 |
| *G* | 0.03063 | 1 | 0.03063 | 11.87 | 13.36 |
| *AB* | 0.01563 | 1 | 0.01563 | 6.06 | 6.82 |
| *AF* | 0.05063 | 1 | 0.05063 | 19.62 | 22.09 |
| *CE* | 0.01563 | 1 | 0.01563 | 6.06 | 6.82 |
| Pooled error | 0.02063 | 8 | 0.00258 | | 8.99 |
| Total | 0.22923 | 15 | | | 100.00 |

**TABLE 12.16** Average response of significant main effects

| *Factor main effect* | *A* | *D* | *F* | *G* |
|---|---|---|---|---|
| Level 1: | 0.405 | **0.502** | 0.427 | **0.505** |
| Level 2: | 0.517 | 0.420 | 0.495 | 0.417 |

**TABLE 12.17** Average response of significant interaction effects

| *Average of AB* | | *Average of AF* | | *Average of CE* | |
|---|---|---|---|---|---|
| $A_1B_1$ | 0.450 | $A_1F_1$ | 0.315 | $C_1E_1$ | 0.455 |
| $A_1B_2$ | 0.360 | $A_1F_2$ | 0.495 | $C_1E_2$ | 0.490 |
| $A_2B_1$ | 0.500 | $A_2F_1$ | **0.540** | $C_2E_1$ | **0.495** |
| $A_2B_2$ | **0.535** | $A_2F_2$ | 0.495 | $C_2E_2$ | 0.405 |

Suppose the objective is to maximize the response. From Table 12.17, the combinations $A_2B_2$, $A_2F_1$ and $C_2E_1$ results in maximum response. So, the optimal levels for $A$, $B$, $C$ and $F$ are 2, 2, 2 and 1 respectively. For $D$ and $G$, the optimal levels are 1 and 1 (Table 12.16) respectively. So, the optimal condition is $A_2$ $B_2$ $C_2$ $D_1$ $F_1$ $G_1$.

## 12.5 ATTRIBUTE DATA ANALYSIS

Variable data is obtained through measurement of the quality characteristic (response). We get attribute data when we classify the product into different categories such as good/bad or accepted/rejected. When the response from the experiment is in the form of attribute data, how to analyse is discussed in this section. Usually, the response will be either defectives or defects. There are two approaches to analyse the data through ANOVA. One approach is to treat the defectives (response) from each experiment as variable data and analyse. In the second approach, we use both the classes of data (defectives and non-defectives) and represent mathematically these two classes as 1 and 0 respectively. Suppose in a sample of 20 parts, 3 are defective and 17 are non-defective (good). This is treated as three 1s and seventeen 0s. These 1s and 0s are considered as data and analysed. These two approaches are discussed using the data from $L_8$ OA given in Table 12.18.

**TABLE 12.18** $L_8$ OA: Attribute data

| *Trial no.* | *Factors/Columns* | | | | | | | *Defective (bad)* | *Non-defective (good)* |
|---|---|---|---|---|---|---|---|---|---|
| | *A* 1 | *B* 2 | *C* 3 | *D* 4 | *E* 5 | *F* 6 | *G* 7 | | |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 5 | 20 |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 7 | 18 |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 23 |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 4 | 21 |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 3 | 22 |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 24 |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 6 | 19 |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 0 | 25 |

### 12.5.1 Treating Defectives as Variable Data

In this approach we consider the defectives (bad) as response from the experiment and ignore the non-defectives. Accordingly, the objective of this experiment will be to minimize the number of defectives. Considering the number of defectives from each experiment as variable, the sum of squares is computed and ANOVA is carried out.

Total number of defectives ($T$) = 28

Total number of observations ($N$) = 8 (treated as one replicate)

$$\text{CF} = \frac{T^2}{N} = \frac{(28)^2}{8} = 98.00$$

$$SS_{\text{Total}} = (5)^2 + 7)^2 + \ldots + (6)^2 + (0)^2 - \text{CF}$$

$$= 140.00 - 98.00 = 42.00$$

The response totals required for computing sum of squares is given in Table 12.19.

**TABLE 12.19** Response totals for the attribute data

| *Factors* | *A* | *B* | *C* | *D* | *E* | *F* | *G* |
|---|---|---|---|---|---|---|---|
| Level 1 | 18 | 16 | 18 | 16 | 8 | 12 | 16 |
| Level 2 | 10 | 12 | 10 | 12 | 20 | 16 | 12 |

Note that there are four observations in each level total.

$$SS_A = \frac{(18)^2}{4} + \frac{(10)^2}{4} - \text{CF}$$

$$= \frac{(18)^2 + (10)^2}{4} - \text{CF}$$

$$= 106.00 - 98.00 = 8.00$$

Similarly, other factor sum of squares is computed.

$$SS_B = 2.00,\ SS_C = 8.00,\ SS_D = 2.00,\ SS_E = 18.00,\ SS_F = 2.00,\ SS_G = 2.00$$

These computations are summarized in the initial ANOVA Table 12.20.

**TABLE 12.20** ANOVA (initial) for the attribute data

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *C*(%) |
|---|---|---|---|
| *A* | 8.00 | 1 | 19.05 |
| *B* | 2.00 | 1 | 4.76 |
| *C* | 8.00 | 1 | 19.05 |
| *D* | 2.00 | 1 | 4.76 |
| *E* | 18.00 | 1 | 42.86 |
| *F* | 2.00 | 1 | 4.76 |
| *G* | 2.00 | 1 | 4.76 |
| Error (pure) | 0.00 | – | – |
| Total | 42.00 | 7 | |

Pooling the sum of squares of $B$, $D$, $F$ and $G$ into the error term, we have the final ANOVA (Table 12.21).

**TABLE 12.21** ANOVA (final) for the attribute data

| *Source of variation* | *Sum of squares* | *Degress of freedom* | *Mean square* | $F_0$ | $C(\%)$ |
|---|---|---|---|---|---|
| $A$ | 8.00 | 1 | 8.00 | 4.00 | 19.05 |
| $C$ | 8.00 | 1 | 8.00 | 4.00 | 19.05 |
| $E$ | 18.00 | 1 | 18.00 | 9.00 | 42.85 |
| Error (pooled) | 8.00 | 4 | 2.00 | | 19.05 |
| Total | 42.00 | 7 | | | 100.00 |

It is observed from ANOVA (Table 12.21) that at 5% level of significance only factor $E$ is significant and contributes 42.9% to the total variation. Though factors $A$ and $C$ are not significant, their contribution seems to be considered as important. Also we use one-half of the degrees of freedom for predicting the optimal value. Therefore, the optimal levels for minimization of response (defectives) is $A_2$, $C_2$ and $E_1$ (Table 12.19).

The predicted optimum response ($\mu_{\text{pred}}$) is given by

$$\mu_{\text{pred}} = \bar{Y} + (\bar{A}_2 - \bar{Y}) + (\bar{C}_2 - \bar{Y}) + (\bar{E}_1 - \bar{Y}) \tag{12.9}$$

$$= \bar{A}_2 + \bar{C}_2 + \bar{E}_1 - 2\bar{Y}$$

$$\text{Overall average fraction defective } (\bar{Y}) = \frac{\text{Total number of defectives}}{\text{Total number of items inspected}} \tag{12.10}$$

$$(\bar{Y}) = \frac{28}{200} = 0.14 \text{ or } 14\%$$

Similarly, $\bar{A}_2 = \dfrac{10}{200} = 0.05$, $\bar{C}_2 = \dfrac{10}{200} = 0.05$ and $\bar{E}_1 = \dfrac{8}{200} = 0.04$

Substituting these values in Eq. (12.9), we get

$$\mu_{\text{pred}} = \bar{A}_2 + \bar{C}_2 + \bar{E}_1 - 2\bar{Y}$$

$$= (0.05 + 0.05 + 0.04) - 2\,(0.14) = -0.14$$

$\mu_{\text{pred}}$ is a percent defective and it cannot be less than zero (negative). This will happen when we deal with fractions or percentage data. How to deal with this is explained in Section 12.5.3.

### 12.5.2 Considering the Two-class Data as 0 and 1

Consider the same data used in Table 12.18. For trial 1, the bad parts are 5 and 20 good parts. The 5 bad products are treated as five 1s and the 20 good parts are treated as twenty 0s. If the data is considered as 1s and 0s, the total sum of squares is given by

$$SS_{\text{Total}} = (1)^2 + (1)^2 + (1)^2 + (1)^2 + (1)^2 + (0)^2 + (0)^2 + \ldots + (0)^2 - \text{CF}$$

$$= T - \frac{T^2}{N} \tag{12.11}$$

where,

$T$ = sum of all data

$N$ = total number of parts inspected = 200

$$SS_{\text{Total}} = 28 - \frac{(28)^2}{200} = \frac{5600 - 784}{200} = \frac{4816}{200} = 24.08$$

The sum of squares of columns (factors) is computed similar to that of variable data. Note that the number of observations in each total here is 100.

$$SS_A = \frac{(18)^2}{100} + \frac{(10)^2}{100} - \text{CF}$$

$$= 4.24 - 3.92$$

$$= 0.32$$

Similarly other sum of squares is computed.

$$SS_B = 0.08,\ SS_C = 0.32,\ SS_D = 0.08,\ SS_E = 0.72,\ SS_F = 0.08,\ SS_G = 0.08$$

and $SS_e = SS_{\text{Total}}$ – SS of all factors

$$= 24.08 - (0.32 + 0.08 + 0.32 + 0.08 + 0.72 + 0.08 + 0.08) = 22.4.$$

These computations are summarized in Table 12.22.

**TABLE 12.22** ANOVA (initial) for the two-class attribute data

| *Source of variation* | *Sum of squares* | *Degress of freedom* | *Mean squares* | $F_0$ |
|---|---|---|---|---|
| *A* | 0.32 | 1 | 0.32 | 2.74 |
| *B* | 0.08 | 1 | 0.08 | 0.68 |
| *C* | 0.32 | 1 | 0.32 | 2.74 |
| *D* | 0.08 | 1 | 0.08 | 0.68 |
| *E* | 0.72 | 1 | 0.72 | 6.15 |
| *F* | 0.08 | 1 | 0.08 | 0.68 |
| *G* | 0.08 | 1 | 0.08 | 0.68 |
| Error (pure) | 22.40 | 192 | 0.117 | |
| Total | 24.08 | 199 | | |

At 5% significance level, only factor *E* is significant. The final ANOVA with the pooled error is given in Table 12.23.

**TABLE 12.23** ANOVA (final) for the two-class attribute data

| *Source of variation* | *Sum of squares* | *Degress of freedom* | *Mean square* | $F_0$ | *C(%)* |
|---|---|---|---|---|---|
| *A* | 0.32 | 1 | 0.32 | 2.76 | 0.84 |
| *C* | 0.32 | 1 | 0.32 | 2.76 | 0.84 |
| *E* | 0.72 | 1 | 0.72 | 6.21 | 2.51 |
| Error (pooled) | 22.72 | 196 | 0.116 | | 95.81 |
| Total | 24.08 | 199 | | | 100.00 |

This result is same as that obtained in Section 12.5.1. In this case, the pooling error variance does not affect the results because the error degrees of freedom are already large. Also, the usual interpretation of percent contribution with this type of data analysis will be misleading.

### 12.5.3 Transformation of Percentage Data

The prediction of optimum response is based on additivity of individual factor average responses. When we deal with percentage (fractional) data such as percent defective or percent yield, the value may approach 0% or 100% respectively. Because of poor additivity of percentage data, the predicted value may be more than 100% or less than 0% (negative). Hence, the data is transformed (Omega transformation) into dB values using Omega tables (Appendix D) or formula. Then the $\mu_{\text{pred}}$ is obtained in terms of dB value. This value is again converted back to percentage using the Omega table or formula.

The Omega transformation formula is: $\Omega(\text{dB}) = 10 \log \dfrac{p}{1-p}$ (12.12)

where, $p$ is the fraction defective ($0 < p < 1$).

As an example, consider the illustration discussed in Section 12.5.1.

$$\bar{Y} = 14\%, \bar{A}_2 = 5\%, \bar{C}_2 = 5\% \text{ and } \bar{E}_1 = 4\%$$

$$\begin{aligned}\mu_{\text{pred}} &= \bar{A}_2 + \bar{C}_2 + \bar{E}_1 - 2\bar{Y} \\ &= (0.05 + 0.05 + 0.04) - 2(0.14) \\ &= -0.14 \text{ or } -14\% \text{ which is not meaningful.}\end{aligned}$$

Applying transformation,

| | | | |
|---|---|---|---|
| Per cent defective: | 4 | 5 | 14 |
| dB value: | –13.801 | –12.787 | –7.883 |

Now substituting these dB values, we obtain

$$\begin{aligned}\mu_{\text{pred}} &= (-12.787 - 12.787 - 13.801) - 2(-7.883) \\ &= -39.375 + 15.766 \\ &= -23.609 \text{ dB}\end{aligned}$$

Converting dB into percent defective, $\mu_{\text{pred}}$ is about 0.4%.

## 12.6 CONFIRMATION EXPERIMENT

The purpose of confirmation experiment is to validate the conclusions drawn from the experiment. Confirmation is necessary, when we use OA designs and fractional factorial designs because of confounding within the columns of OA and the presence of aliases respectively. Confirmation experiment is conducted using the optimal levels of the significant factors. For the insignificant factors, although any level can be used, usually levels are selected based on economics and convenience. The sample size for the confirmation experiment is larger than the sample size of any one experimental trial in the original experiment.

The mean value estimated from the confirmation experiment ($\mu_{\text{conf}}$) is compared with $\mu_{\text{pred}}$ for validating the experiment/results. Generally if $\mu_{\text{conf}}$ is within ± 5% of $\mu_{\text{pred}}$, we can assume a good agreement between these two values. A good agreement between $\mu_{\text{conf}}$ and $\mu_{\text{pred}}$ indicate that additivity is present in the model and interaction effects can not be dominant. Poor agreement between $\mu_{\text{conf}}$ and $\mu_{\text{pred}}$ indicate that additivity is not present and there will be poor reproducibility of small scale experiments to large scale production and the experimenter should not implement the predicted optimum condition on a large scale.

## 12.7 CONFIDENCE INTERVALS

The predicted mean ($\mu_{\text{pred}}$ ) from the experiment as well as the mean estimated from the confirmation experiment are point estimates based on the averages of results. Statistically, this provides a 50% chance that the true average may be greater and 50% chance that the true average may be less than $\mu_{\text{pred}}$. The experimenter would like to have a range of values within which the true average be expected to fall with some confidence. This is called the *confidence interval*.

### 12.7.1 Confidence Interval for a Treatment Mean

$$\text{CI}_1 = \sqrt{\frac{F_{\alpha,\nu_1,\nu_2} \times MS_e}{n}} \tag{12.13}$$

where

$F_{\alpha,\nu_1,\nu_2}$ = $F$-value from table
$\alpha$ = significance level
$\nu_1$ = the numerator degrees of freedom associated with mean which is always 1
$\nu_2$ = degrees of freedom for pooled error mean square
$MS_e$ = pooled error mean square/variance
$n$ = number of observations used to calculate the mean

Suppose we want to compute confidence interval for the factor level/treatment, $\overline{B}_1$ (Illustration 12.2.).

$$\mu_{B_1} = \overline{B}_1 \pm \text{CI}_1$$

$$\overline{B}_1 - \text{CI}_1 \le \mu_{B_1} \le \overline{B}_1 + \text{CI}_1$$

At $\alpha$ = 5%, $F_{0.05,1,19}$ = 4.38

$$n = 12,\ \bar{B}_1 = 7.08 \text{ and } MS_e = 12.92$$

$$\text{CI}_1 = \sqrt{\frac{4.38 \times 12.92}{12}} = 2.17$$

The confidence interval range is

$$7.08 - 2.17 \le \mu_{B1} \le 7.08 + 2.17$$

$$4.91 \le \mu_{B1} \le 9.25$$

### 12.7.2 Confidence Interval for Predicted Mean

The confidence interval for the predicted mean ($\mu_{\text{pred}}$) is given by

$$\text{CI}_2 = \sqrt{\frac{F_{\alpha,\nu_1,\nu_2} \times MS_e}{n_{\text{eff}}}} \tag{12.14}$$

where $n_{\text{eff}}$ is effective number of observations.

$$n_{\text{eff}} = \frac{\text{Total number of experiments}}{1 + \text{sum of df of effects used in } \mu_{\text{pred}}} \tag{12.15}$$

From Illustration 12.2,

$$\mu_{\text{pred}} = \bar{B}_1 + \bar{E}_2 + \overline{A_1D_1} - \bar{D}_1 - \bar{Y} = 9.87 \quad \text{[Eq. (12.7)]}$$

Total number of experiments = 8 × 3 = 24 and there are four effects in $\mu_{\text{pred}}$.

Therefore, $$n_{\text{eff}} = \left(\frac{24}{1+4}\right) = \left(\frac{24}{5}\right) = 4.8$$

$$\text{CI}_2 = \sqrt{\frac{F_{\alpha,\nu_1,\nu_2} \times MS_e}{n_{\text{eff}}}} = \sqrt{\frac{4.38 \times 12.92}{4.8}} = 3.43$$

The confidence interval for the predicted mean ($\mu_{\text{pred}}$) is given by

$$\mu_{\text{pred}} - \text{CI}_2 \le \mu_{\text{pred}} \le \mu_{\text{pred}} + \text{CI}_2$$

$$(9.87 - 3.43) \le \mu_{\text{pred}} \le (9.87 + 3.43)$$

$$6.44 \le \mu_{\text{pred}} \le 13.30$$

### 12.7.3 Confidence Interval for Confirmation Experiment

$$\text{CI}_3 = \sqrt{F_{\alpha,\nu_1,\nu_2} \times MS_e\left(\frac{1}{n_{\text{eff}}} + \frac{1}{r}\right)} \tag{12.16}$$

where $r$ is the sample size for the confirmation experiment.

Suppose for Illustration 12.2, a confirmation experiment is run 10 times and the mean value is 9.52. The confidence interval is

$$\text{CI}_3 = \sqrt{F_{0.05,1,19} \times 12.92\left(\frac{1}{4.8} + \frac{1}{10}\right)}$$

$$= \sqrt{4.38 \times 12.92\left(\frac{1}{4.8} + \frac{1}{10}\right)} = 4.17$$

Therefore, the confidence interval range is

$$\mu_{\text{conf}} - \text{CI}_3 \le \mu_{\text{conf}} \le \mu_{\text{conf}} + \text{CI}_3$$

$$9.52 - 4.17 \le \mu_{\text{conf}} \le 9.52 + 4.17$$

$$5.35 \le \mu_{\text{conf}} \le 13.69$$

The confidence interval for the predicted mean ($\mu_{\text{pred}}$) and the confidence interval for the confirmation experiment ($\mu_{\text{conf}}$) are compared to judge whether the results are reproducible. If the confidence interval range of the predicted mean ($\mu_{\text{pred}} \pm \text{CI}_2$ ) overlaps with the confidence interval range of the confirmation experiment ($\mu_{\text{conf}} \pm \text{CI}_3$), we may accept that the results are additive and the experimental results are reproducible.

For example, Illustration 12.2, the confidence interval of the predicted mean (6.44 to 13.30) overlaps fairly well with the confidence interval of the confirmation experiment (5.35 to 13.69). Thus, we may infer that the experimental results are reproducible.

## PROBLEMS

**12.1** An industrial engineer has conducted an experiment on a welding process. The following control factors have been investigated (Table 12.24a).

**TABLE 12.24a** Data for Problem 12.1

| *Factors* | *Level* 1 | *Level* 2 |
|---|---|---|
| Weld material (*A*) | SS 41 | SB 35 |
| Type of welding rod (*B*) | J100 | B17 |
| Thickness of weld metal (*C*) | 8 mm | 12 mm |
| Weld current (*D*) | 130 A | 150 A |
| Preheating (*E*) | No | heating by 100°C |
| Opening of welding parts (*F*) | 2 mm | 3 mm |
| Drying of welding rod (*G*) | No drying | 2 days drying |

The experiment was conducted using $L_8$ OA and the data (weld strength in kN) obtained are given in Table 12.24b.

**TABLE 12.24b** Data for Problem 12.1

| *Trial no.* | *Factors/Columns* | | | | | | | *Results* | |
|---|---|---|---|---|---|---|---|---|---|
| | *A* | *B* | *C* | *D* | *E* | *F* | *G* | $R_1$ | $R_2$ |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 20 | 17 |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 22 | 28 |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 39 | 34 |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 27 | 37 |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 23 | 30 |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 10 | 6 |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 8 | 7 |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 17 | 23 |

(a) Determine the average response for each factor level and identify the significant effects.
(b) What is the predicted weld strength at the optimum condition?

**12.2** Consider the data given in Problem 12.1 and analyse using ANOVA and interpret the results.

**12.3** The data is available from an experimental study (Table 12.25).

**TABLE 12.25** Data for Problem 12.3

| *Trial no.* | *Factors/Columns* | | | | | | | *Results* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *A* | *B* | *AB* | *C* | *AC* | *D* | *E* | $R_1$ | $R_2$ | $R_3$ |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | | |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 20 | 19 | 17 |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 22 | 24 | 28 |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 39 | 35 | 34 |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 27 | 26 | 37 |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 23 | 25 | 30 |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 10 | 12 | 6 |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 8 | 6 | 7 |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 17 | 18 | 23 |

(a) Analyse the data using response graph method and comment on the results.
(b) Analyse the data using ANOVA and comment on the results.

**12.4** An experiment was conducted with eight factors each at three levels on the flash butt welding process. The weld joints were tested for welding defects and the following data were obtained (Table 12.26). Analyse the data and determine the optimal levels for the factors.

**TABLE 12.26** Data for Problem 12.4

| *Trial no.* | *Factors/Columns* | | | | | | | | *Response* | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *A* 1 | *B* 2 | *C* 3 | *D* 4 | *E* 5 | *F* 6 | *G* 7 | *H* 8 | *Good* | *Bad* |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 6 | 0 |
| 2 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 6 | 0 |
| 3 | 1 | 1 | 3 | 3 | 3 | 3 | 3 | 3 | 5 | 1 |
| 4 | 1 | 2 | 1 | 1 | 2 | 2 | 3 | 3 | 6 | 0 |
| 5 | 1 | 2 | 2 | 2 | 3 | 3 | 1 | 1 | 6 | 0 |
| 6 | 1 | 2 | 3 | 3 | 1 | 1 | 2 | 2 | 6 | 0 |
| 7 | 1 | 3 | 1 | 2 | 1 | 3 | 2 | 3 | 6 | 0 |
| 8 | 1 | 3 | 2 | 3 | 2 | 1 | 3 | 1 | 6 | 0 |
| 9 | 1 | 3 | 3 | 1 | 3 | 2 | 1 | 2 | 6 | 0 |
| 10 | 2 | 1 | 1 | 3 | 3 | 2 | 2 | 1 | 3 | 3 |
| 11 | 2 | 1 | 2 | 1 | 1 | 3 | 3 | 2 | 5 | 1 |
| 12 | 2 | 1 | 3 | 2 | 2 | 1 | 1 | 3 | 4 | 2 |
| 13 | 2 | 2 | 1 | 2 | 3 | 1 | 3 | 2 | 6 | 0 |
| 14 | 2 | 2 | 2 | 3 | 1 | 2 | 1 | 3 | 6 | 0 |
| 15 | 2 | 2 | 3 | 1 | 2 | 3 | 2 | 1 | 6 | 0 |
| 16 | 2 | 3 | 1 | 3 | 2 | 3 | 1 | 2 | 6 | 0 |
| 17 | 2 | 3 | 2 | 1 | 3 | 1 | 2 | 3 | 5 | 1 |
| 18 | 2 | 3 | 3 | 2 | 1 | 2 | 3 | 1 | 5 | 1 |

**12.5** A study was conducted involving three factors *A*, *B* and *C*, each at three levels. The data is given in Table 12.27. Analyse using ANOVA and identify optimal levels for the factors. The objective is to minimize the response.

**TABLE 12.27** Data for Problem 12.5

| *Trial no.* | *Factors/Columns* | | | | *Results* | |
|---|---|---|---|---|---|---|
| | *A* 1 | *B* 2 | *C* 3 | *D* 4 | $R_1$ | $R_2$ |
| 1 | 1 | 1 | 1 | 1 | 2.32 | 2.50 |
| 2 | 1 | 2 | 2 | 2 | 2.40 | 2.80 |
| 3 | 1 | 3 | 3 | 3 | 1.75 | 2.80 |
| 4 | 2 | 1 | 2 | 3 | 0.85 | 1.00 |
| 5 | 2 | 2 | 3 | 1 | 1.70 | 1.50 |
| 6 | 2 | 3 | 1 | 2 | 1.30 | 1.60 |
| 7 | 3 | 1 | 3 | 2 | 2.35 | 1.30 |
| 8 | 3 | 2 | 1 | 3 | 1.60 | 2.30 |
| 9 | 3 | 3 | 2 | 1 | 1.40 | 1.10 |

CHAPTER 13

# Robust Design

## 13.1 INTRODUCTION

Robust design is an engineering methodology used to improve productivity during Research and Development (R&D) activities so that high quality products can be developed fast and at low cost. We say that a product or a process is robust if its performance is not affected by the noise factors. And robust design is a procedure used to design products and processes such that their performance is insensitive to noise factors. In this, we try to determine product parameters or process factor levels so as to optimize the functional characteristics of products and have minimal sensitivity to noise. Robust design was developed by Dr. Genich Taguchi in the 1950s, generally called as *Taguchi methods*. Orthogonal Array experiments are used in *Robust design*.

Traditional techniques of analysis of experimental designs focus on identifying factors that affect mean response. These are termed *location effects* of the factors. In the technique of ANOVA, it is assumed that the variability in the response characteristic remains more or less the same from one trial to another. And on this assumption of equality of variances for the treatment combinations, the mean responses are compared. When this assumption is violated, the results will be inaccurate. To overcome this problem, statisticians have suggested that the data be suitably transformed before performing ANOVA.

Taguchi was the first to suggest that statistically planned experiments should be used in the product development stage to detect factors that affect variability of the output termed *dispersion effects* of the factors. He argued that by setting the factors with important dispersion effects at their optimal levels, the output can be made robust to changes in the operating and environmental conditions during production. Thus, the identification of dispersion effect is important to improve the quality of a process. In order to achieve a robust product/process one has to consider both location effect and dispersion effect. Taguchi has suggested a combined measure of both these effects. Suppose $m$ is the mean effect and $\sigma^2$ represent variance (dispersion effect). These two measures are combined into a single measure represented by $m^2/\sigma^2$. In terms of communications engineering terminology $m^2$ may be termed as the power of the signal and $\sigma^2$ may be termed as the power of noise and is called the SIGNAL TO NOISE RATIO (S/N ratio). The data is transformed into S/N ratio and analysed using ANOVA and then optimal levels for the factors is determined. This leads to the development of a robust process/product. In Taguchi methods, the word optimization means determination of best levels for the control factors that minimizes the effect of noise. The best levels of control factors are those that maximize the S/N ratio.

Usually, in robust design only main factors are studied along with the noise factors. Interactions are not included in the design.

## 13.2 FACTORS AFFECTING RESPONSE

A number of factors influence the performance (response) of a product or process. These factors can be classified into the following:

- Control factors
- Noise factors
- Signal factors
- Scaling factors

***Control factors:*** These factors are those whose values remain fixed once they are chosen. These factors can be controlled by the manufacturer and cannot be directly changed by the customer. These include design parameters (specifications) of a product or process such as product dimensions, material, configuration etc.

***Noise factors:*** These factors are those over which the manufacturer does not have any control and they vary with the product's usage and its environment. These noise factors are further classified as follows:

*Outer noise:* Produces variation from outside the product/process (environmental factors).
*Inner noise:* Produces variation from inside or within the product/process (functional and time related).
*Product noise:* Part to part variation.

Examples for these noise factors related to a product and a process are listed in Table 13.1.

**TABLE 13.1** Examples of noise factors

| *Product* | *Process* |
|---|---|
| *Outer Noise* | |
| Customer's usage | Incoming material |
| Temperature | Temperature |
| Humidity | Humidity |
| Dust | Dust |
| Solar radiation | Voltage variation |
| Shock | Operator performance |
| Vibration | Batch to batch variation |
| *Inner Noise* | |
| Deterioration of parts | Machinery aging |
| Deterioration of material | Tool wear |
| Oxidation | Process shift in control |
| *Between Products* | |
| Piece to piece variation | Process to process variation |

***Signal factors:*** These factors change the magnitude of the response variable. In static functions (problems) where the output (response) is a fixed target value (maximum strength, minimum wear or a nominal value), the signal factor takes a constant value. Here, the optimization problem involves the determination of the best levels for the control factors so that the output is at the target value. Of course, the aim is to minimize the variation in output even though noise is present.

In dynamic problems, where the quality characteristic (response) will have a variable target value, the signal factor also varies. In many applications, the output follow input signal in a predetermined manner. For example, accelerator peddles in cars ( application of pressure), volume control in audio amplifiers (angle of turn of knob), etc. In these problems, we need to evaluate the control and signal factors in the presence of noise factors.

***Scaling factors:*** These factors are used to shift the mean level of a quality characteristic to achieve the required functional relationship between the signal factor and the quality characteristic (adjusting the mean to the target value). Hence, this factor is also called adjustment factor.

## 13.3 OBJECTIVE FUNCTIONS IN ROBUST DESIGN

Taguchi extended the audio concept of the signal to noise ratio to multi-factor experiments. The signal to noise ratio (S/N ratio) is a statistic that combines the mean and variance. The objective in robust design is to minimize the sensitivity of a quality characteristic to noise factors. This is achieved by selecting the factor levels corresponding to the maximum S/N ratio. That is, in setting parameter levels we always maximize the S/N ratio irrespective of the type response (i.e., maximization or minimization). These S/N ratios are often called *objective functions* in robust design.

### *Types of S/N ratios*

The static problems associated with batch process optimization are common in the industry. Where as dynamic problems are associated with technology development situations. Hence, only S/N ratios used for optimization of static problems are discussed in this chapter. The S/N ratio depends on the type of quality characteristic. Often we deal with the following four types of quality characteristics:

- Smaller–the better
- Nominal–the best
- Larger–the better
- Fraction defective

***Smaller–the better:*** Here, the quality characteristic is continuous and non-negative. It can take any value between 0 – ∞. The desired value (the target) is zero. These problems are characterized by the absence of scaling factor (ex: surface roughness, pollution, tyre wear, etc.). The S/N ratio ($\eta$) is given by

$$\eta = -10 \log\left[\frac{1}{n} \sum_{i=1}^{n} Y_i^2\right] \tag{13.1}$$

where $n$ is the number of replications.

***Nominal–the best:*** In these problems, the quality characteristic is continuous and non-negative. It can take any value from 0 to ∞. Its target value is non-zero and finite. We can use adjustment factor to move mean to target in these types of problems.

$$\text{The S/N ratio } (\eta) = 10 \log \left( \frac{\bar{Y}^2}{S^2} \right) \tag{13.2}$$

where

$$\bar{Y} = \sum_{i=1}^{n} \frac{Y_i}{n} \text{ and}$$

$$S^2 = \sum_{i=1}^{n} \frac{(Y_i - \bar{Y})^2}{n} = \frac{\sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2}{n}$$

The optimization is a two step process. First we select control factor levels corresponding to maximum $\eta$, which minimizes noise. In the second step, we identify the adjustment factor and adjust its mean to the target value. Adjustment factor can be identified after experimentation through ANOVA or from engineering knowledge/experience of the problem concerned. The adjustment factor can be one of those factors which affect mean only (significant in raw data ANOVA) and no effect on $\eta$. Sometimes we consider a factor that has a small effect on $\eta$.

***Larger–the better:*** The quality characteristic is continuous and non-negative. It can take any value from 0 to ∞.

The ideal target value of this type quality characteristic is ∞ (as large as possible). Quality characteristics like strength values, fuel efficiency, etc. are examples of this type. In these problems, there is no scaling factor. The S/N ratio ($\eta$) is given by

$$\eta = -10 \log \left( \frac{1}{n} \sum_{i=1}^{n} \frac{1}{Y_i^2} \right) \tag{13.3}$$

***Fraction defective:*** When the quality characteristic is a proportion ($p$) which can take a value from 0 to 1.0 (0 to 100%), the best value of $p$ is zero in the case of fraction defective or 1.0 if it is percentage of yield.

The objective function here is

$$\eta = -10 \log \left( \frac{1}{p - 1} \right) \quad \text{or} \quad \eta = 10 \log \left( \frac{p}{1 - p} \right) \tag{13.4}$$

where $p$ is the fraction defective.

It is called *Omega transformation*. The range of values for $p$ is 0 to 1.0 and that of $\eta$ is $-\infty$ to $+\infty$.

## 13.4 ADVANTAGES OF ROBUST DESIGN

The advantages of robust design are as follows:

- Dampen the effect of noise (reduce variance) on the performance of a product or process by choosing the proper levels for the control factors. That is, make the product/process robust against the noise factors.
- Improve quality (reduce variation) without removing the cause of variation. The example below explains this.

A tile manufacturing company had a high percentage of rejection. When investigated, it was found that the tiles in the mid zone of the kiln used for firing did not get enough and uniform heat (temperature). This was the cause for rejection. Experts suggested two alternatives to eliminate/reduce the rejection percentage.

(i) Redesign the kiln to remove the cause. This requires heavy expenditure.
(ii) Conduct a robust design experiment.

They have selected the second alternative and conducted an experiment and determined the optimal levels for the process parameters and implemented. The rejections were almost zero. So, the cause is not removed but quality has improved due to reduced variation.

## 13.5 SIMPLE PARAMETER DESIGN

The simplest form of parameter design (Robust design) is one in which the noise factors are treated like control factors and data are collected. The control factors are assigned to the columns of OA and each experimental trial is repeated for each level of noise factor and data are collected. This type of design is often used when we have only one or two noise factors.

### *Simple parameter design with one noise factor*

The design is illustrated with an experiment. An experiment was conducted with an objective of increasing the hardness of a die cast engine component. Also uniform surface hardness (consistency of hardness from position to position) was desired. A total of seven factors each at two levels were studied using $L_8$ OA. Table 13.2 gives this design. In this design, position is the noise factor. At each position, one observation is obtained for each trial. After the data is collected, for each trial, the S/N ratio is computed using the appropriate objective function. This S/N ratio is treated as response and analysed using either response graph method or ANOVA as discussed in Chapter 12.

### *Simple parameter design with two noise factors*

This design is explained with a problem. The problem is concerned with a heat treatment process of a manufactured part. The following control factors and noise factors have been identified for investigation. All the factors are to be studied at two levels. Since it is difficult to control the transfer time and quenching oil temperature during production, these two are considered as noise factors. Apart from the main effects, the interactions $AB$ and $BC$ are also suspected to have an influence on the response (hardness). The required design to study this problem is given in Table 13.3. The data analysis can be carried out after computing S/N ratio for each trial.

**TABLE 13.2** Example of a simple parameter design with one noise factor

| *Trial no.* | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *Position* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | $P_1$ | $P_2$ | $P_3$ | $P_4$ |
| | | | | | | | | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | * | * | * | * |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | * | * | * | * |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | * | * | * | * |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | * | * | * | * |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | * | * | * | * |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | * | * | * | * |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | * | * | * | * |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | * | * | * | * |

**TABLE 13.3** Example of a simple parameter design with two noise factors

| *Trial no.* | *A* | *B* | *AB* | *C* | *D* | *BC* | *E* | *Response (hardness)* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | $T_1$ | | $T_2$ | |
| | | | | | | | | $t_1$ | $t_2$ | $t_1$ | $t_2$ |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | * | * | * | * |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | * | * | * | * |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | * | * | * | * |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | * | * | * | * |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | * | * | * | * |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | * | * | * | * |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | * | * | * | * |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | * | * | * | * |

*Control factors:* Heating temperature (*A*), Heating time (*B*), Quenching duration (*C*), Quenching method (*D*), Quenching media (*E*).

*Noise factors:* Transfer time (*T*), Quenching oil temperature (*t*).

**ILLUSTRATION 13.1**

**Simple Parameter Design with One Noise Factor**

A study was conducted on a manual arc welding process. The objective was to maximize the welding strength. The following control factors each at two levels were studied. In addition to the control factors the interactions *AD* and *CD* were also considered.

| *Factors* | *Level* 1 | *Level* 2 |
|---|---|---|
| Weld design (*A*) | Existing | Modified |
| Cleaning method (*B*) | Wire brush | Grinding |
| Preheating temperature (*C*) | 100°C | 150°C |
| Post weld heat treatment (*D*) | Done | Not done |
| Welding current (*E*) | 40 A | 50 A |

Each trial was repeated by three operators ($O$). Since variability due to operators affect weld quality, operator was treated as noise factor. That is, the noise factor has three levels. A simple parameter design with one noise factor was used and data was collected on the weld strength in kilo Newton (kN). The design with data is given in Table 13.4.

**TABLE 13.4** Parameter design with one noise factor

| *Trial no.* | *D* | *C* | *CD* | *A* | *AD* | *B* | *E* | *Operator (O)* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 1 | 2 | 3 |
| | | | | | | | | $Y_1$ | $Y_2$ | $Y_3$ |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 31 | 24 | 31 |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 24 | 24 | 24 |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 24 | 21 | 34 |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 24 | 20 | 28 |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 29 | 28 | 24 |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 24 | 21 | 21 |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 21 | 24 | 24 |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 34 | 24 | 28 |
| | | | | | | | Total | 211 | 186 | 214 |

### *Data analysis using ANOVA*

The operator-wise total response for all the main factor effects is given in Table 13.5. The operator-wise total response for the interactions *AD* and *CD* are given in Tables 13.6 and 13.7 respectively.

**TABLE 13.5** Operator-wise response totals for the main effects

| *Main effect* | *Level* | *Operator (O)* | | | *Total* |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | |
| *A* | 1 | 105 | 97 | 113 | 315 |
| | 2 | 106 | 89 | 101 | 296 |
| *B* | 1 | 118 | 96 | 111 | 325 |
| | 2 | 93 | 90 | 103 | 286 |
| *C* | 1 | 108 | 97 | 100 | 305 |
| | 2 | 103 | 89 | 114 | 306 |
| *D* | 1 | 103 | 89 | 117 | 309 |
| | 2 | 108 | 97 | 97 | 302 |
| *E* | 1 | 100 | 89 | 104 | 293 |
| | 2 | 111 | 97 | 110 | 318 |

**TABLE 13.6** Operator-wise total response for the interaction $AD$

| | $A_1$ | | | Total | $A_2$ | | | Total |
|---|---|---|---|---|---|---|---|---|
| | $O_1$ | $O_2$ | $O_3$ | | $O_1$ | $O_2$ | $O_3$ | |
| $D_1$ | 55 | 45 | 65 | 165 | 48 | 44 | 52 | 144 |
| $D_2$ | 50 | 52 | 48 | 150 | 58 | 45 | 49 | 152 |

**TABLE 13.7** Operator-wise total response for the interaction $CD$

| | $C_1$ | | | Total | $C_2$ | | | Total |
|---|---|---|---|---|---|---|---|---|
| | $O_1$ | $O_2$ | $O_3$ | | $O_1$ | $O_2$ | $O_3$ | |
| $D_1$ | 55 | 48 | 55 | 158 | 48 | 41 | 62 | 151 |
| $D_2$ | 53 | 49 | 45 | 147 | 55 | 48 | 52 | 155 |

***Computation of sum of squares:***

Grand total = 611
Total number of observations = 24
Grand mean ($\bar{Y}$) = 611/24 = 25.46
Correction factor (CF) = $(611)^2/24$ = 15,555.04

Total sum of squares ($SS_{\text{Total}}$) = $(31)^2 + (24)^2 + (31)^2 + (24)^2 + \ldots + (28)^2 - \text{CF}$

$$= 15927.00 - 15555.04 = 371.96$$

$$SS_A = \frac{(315)^2}{12} + \frac{(296)^2}{12} - \text{CF} = 15.04$$

Similarly, we obtain $SS_B = 63.67$, $SS_C = 0.04$, $SS_D = 2.04$, $SS_E = 26.04$.

The factor interaction sum of squares ($AD$) are obtained using $AD_1$ and $AD_2$ totals obtained from $AD$ column.

$$SS_{AD} = \frac{(317)^2 + (294)^2}{12} - \text{CF} = 22.04$$

and

$$SS_{CD} = \frac{(313)^2 + (298)^2}{12} - \text{CF} = 9.38$$

Sum of squares due to operator is computed using the three operator totals.

$$SS_O = \frac{(211)^2}{8} + \frac{(186)^2}{8} + \frac{(214)^2}{8} - \text{CF} = 59.08$$

The interaction sum of squares between the factor effects and the noise factor are computed using factor-operator totals.

$$SS_{AO} = \frac{(105)^2 + (97)^2 + (113)^2 + (106)^2 + (89)^2 + (101)^2}{4} - \text{CF} - SS_A - SS_O$$

$$= 15640.25 - 15555.04 - 15.04 - 59.08 = 11.09$$

Similarly, $SS_{BO}$ = 27.26, $SS_{CO}$ = 35.59, $SS_{DO}$ = 59.09, $SS_{EO}$ = 1.59

The interaction between the factor interaction ($AD$ and $CD$) and noise factor ($O$) is computed using the interaction-operator totals (Tables 13.6 and 13.7).

$$SS_{ADO} = \frac{(55)^2 + (50)^2 + (45)^2 + \dots + (52)^2 + (49)^2}{2} - \text{CF} - SS_A - SS_D - SS_O - SS_{AD} - SS_{AO} - SS_{DO}$$

$$= 15758.50 - 15555.04 - 15.04 - 2.04 - 59.08 - 22.04 - 11.09 - 59.09$$

$$= 35.08$$

Similarly, $SS_{CDO}$ = 5.25.

These computations are summarized in Table 13.8.

**TABLE 13.8** ANOVA for Illustration 13.1 (Welding strength study)

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ | *C(%)* |
|---|---|---|---|---|---|
| *A* | 15.04 | 1 | 15.04 | 4.61 | 4.04 |
| *B* | 63.37 | 1 | 63.37 | 19.44** | 17.04 |
| *C** | 0.04 | 1 | 0.04 | — | 0.00 |
| *D** | 2.04 | 1 | 2.04 | — | 0.55 |
| *E* | 26.04 | 1 | 26.04 | 7.99** | 7.00 |
| *AD* | 22.04 | 1 | 22.04 | 6.76** | 5.93 |
| *CD** | 9.37 | 1 | 9.37 | 2.87 | 2.52 |
| *O* | 59.09 | 2 | 29.54 | 9.06** | 15.89 |
| *AO** | 11.09 | 2 | 5.54 | — | 2.98 |
| *BO* | 27.26 | 2 | 13.63 | 4.18 | 7.33 |
| *CO* | 35.59 | 2 | 17.79 | 5.46** | 9.56 |
| *DO* | 59.09 | 2 | 29.54 | 9.06** | 15.89 |
| *EO** | 1.59 | 2 | 0.79 | — | 0.43 |
| *ADO* | 35.08 | 2 | 17.54 | 5.38** | 9.43 |
| *CDO** | 5.25 | 2 | 2.62 | — | 1.41 |
| Pooled error | 29.38 | 9 | 3.26 | 7.90 | |
| Total | 371.96 | 23 | | | 100.00 |

*Pooled into error term $F_{0.05,1,9}$ = 5.12; $F_{0.05,2,9}$ = 4.26, **Significant at 5% level.

From Table 13.8, it can be seen that the difference between operators is significant. And the interaction effects $CO$, $DO$ and $ADO$ are significant even though the main effects $A$, $C$ and $D$ are not significant. Since the operator effect is significant, we will select the optimal factor levels for $A$ and $D$ by considering the average response of $AD$ combinations (Table 13.6). The optimal levels for $B$, $C$ and $E$ are selected based on the average response of these factors. The average response of all the significant effects is given in Table 13.9.

**TABLE 13.9** Average response of the significant effects

| *Effect* | *average response* | *Effect* | *average response* | *Effect* | *average response* |
|---|---|---|---|---|---|
| $A_1D_1$ | **27.5** | $C_1$ | 25.4 | $B_1$ | **27.1** |
| $A_1D_2$ | 25.0 | $C_2$ | **25.5** | $B_2$ | 23.8 |
| $A_2D_1$ | 24.0 | $D_1$ | 25.8 | $E_1$ | 24.4 |
| $A_2D_2$ | 25.3 | $D_2$ | 25.2 | $E_2$ | **26.5** |

The objective of the study is to maximize the weld strength. Therefore, the optimum process parameter combination is $A_1\ B_1\ C_2\ D_1\ E_2$. Note that the average response of factor $C$ at its two levels (25.4 and 25.5) is almost same. Hence, the level for factor $C$ can be 1 or 2.

### *Analysis using S/N data*

The raw data from each trial can be transformed into S/N ratio. The quality characteristic under consideration is larger—the better type (maximization of weld strength). The corresponding objective function [Eq. 13.3)] is

$$\eta = -10 \log \left( \frac{1}{n} \sum_{i=1}^{n} \frac{1}{Y_i^2} \right)$$

Substituting the raw data of trial 1, $\eta = -10 \log \left[ \frac{1}{3} \left( \sum \frac{1}{(31)^2} + \frac{1}{(24)^2} + \frac{1}{(31)^2} \right) \right] = 28.95$

Similarly, for all the trials the S/N data is computed and given in Table 13.10.

**TABLE 13.10** S/N data for Illustration 13.1

| *Trial no.* | *D* | *C* | *CD* | *A* | *AD* | *B* | *E* | *Operator (O)* | | | *S/N ratio* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 1 | 2 | 3 | ($\eta$) |
| | | | | | | | | $Y_1$ | $Y_2$ | $Y_3$ | |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 31 | 24 | 31 | 28.95 |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 24 | 24 | 24 | 27.60 |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 24 | 21 | 34 | 27.90 |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 24 | 20 | 28 | 27.36 |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 29 | 28 | 24 | 28.54 |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 24 | 21 | 21 | 26.80 |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 21 | 24 | 24 | 27.18 |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 34 | 24 | 28 | 28.89 |

Now treating the S/N ratio as response, the S/N data can be analysed using either response graph method or the ANOVA method. We use the ANOVA method so that the result can be compared with the one obtained by the analysis of raw data.

### *S/N data analysis using ANOVA*

The response (S/N) totals for all the factor effects is given in Table 13.11.

**TABLE 13.11** Response (S/N) totals

| *Factor effects* | *Level* 1 | *Level* 2 |
|---|---|---|
| *A* | 112.57 | 110.65 |
| *B* | 113.74 | 109.48 |
| *C* | 111.89 | 111.33 |
| *D* | 111.81 | 111.41 |
| *E* | 110.29 | 112.93 |
| *AD* | 112.54 | 110.68 |
| *CD* | 112.62 | 110.60 |

***Computation of sum of squares:***

Grand total = 223.22

Total number of observations = 8

Grand mean ($\bar{Y}$) = 223.22/8 = 27.9025

Correction Factor (CF) = $(223.22)^2/8$ = 6228.3961

$$\text{Total sum of squares } (SS_{\text{Total}}) = (28.95)^2 + (27.60)^2 + (27.90)^2 + \ldots + (28.89)^2 - \text{CF}$$
$$= 6232.9982 - 6228.3961$$
$$= 4.6021$$

$$SS_A = (112.57)^2 + (110.65)^2/4 - \text{CF}$$
$$= 6228.8569 - 6228.3961$$
$$= 0.4608$$

Similarly, $SS_B$ = 2.2685, $SS_C$ = 0.0392, $SS_D$ = 0.0200, $SS_E$ = 0.8712, $SS_{AD}$ = 0.4324, $SS_{CD}$ = 0.5100

The ANOVA for the S/N data is given in Table 13.12.

**TABLE 13.12** ANOVA for S/N data for Illustration 13.1

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Means square* | $F_0$ | *C(%)* |
|---|---|---|---|---|---|
| *A* | 0.4608 | 1 | 0.4608 | 15.56 | 10.01 |
| *B* | 2.2685 | 1 | 2.2685 | 76.64 | 49.29 |
| *C** | 0.0392 | 1 | — | — | 0.85 |
| *D** | 0.0200 | 1 | — | — | 0.43 |
| *E* | 0.8712 | 1 | 0.8712 | 29.43 | 18.94 |
| *AD* | 0.4324 | 1 | 0.4324 | 14.61 | 9.40 |
| *CD* | 0.5100 | 1 | 0.5100 | 17.23 | 11.08 |
| Pooled error | 0.0592 | 2 | 0.0296 | | |
| Total | 4.6021 | 7 | | | 100.00 |

*Pooled into error

Note that the two effects, whose contribution is less are pooled into the error term and other effects are tested. At 5% level of significance, it is found that *A*, *B*, *E*, *AD* and *CD* are significant. Since the two interactions *AD* and *CD* are significant, the optimal factor levels for *A*, *C* and *D* should be selected based on the average response of the interaction combinations. The average response (S/N) is given in Table 13.13.

**TABLE 13.13** Average response (S/N) of the significant effects

| *Effect* | *average response* | *Effect* | *average response* | *Effect* | *average response* | *Effect* | *average response* |
|---|---|---|---|---|---|---|---|
| $A_1D_1$ | **28.43** | $C_1D_1$ | **28.28** | $A_1$ | **28.14** | $C_1$ | **27.97** |
| $A_1D_2$ | 27.86 | $C_1D_2$ | 27.67 | $A_2$ | 27.66 | $C_2$ | 27.83 |
| $A_2D_1$ | 27.48 | $C_2D_1$ | 27.63 | $B_1$ | **28.44** | $E_1$ | 27.57 |
| $A_2D_2$ | 27.85 | $C_2D_2$ | 28.04 | $B_2$ | 27.37 | $E_2$ | **27.65** |

We know that the optimal levels for factors (based on S/N data) are always selected corresponding to the maximum average response. Accordingly from Table 13.13, the optimum process parameter combination is $A_1$ $B_1$ $C_1$ $D_1$ $E_2$. This result is same as that obtained with the raw (original) data.

## 13.6 INNER/OUTER OA PARAMETER DESIGN

When we have more than two noise factors, we normally use Inner/Outer OA parameter design. Suppose we want to study 7 control factors (*A*, *B*, *C*, *D*, *E*, *F* and *G*), each at two levels and three noise factors (*X*, *Y* and *Z*) each at two levels. For designing an experiment for this problem, we use one OA for control factors, called an Inner OA and another OA for noise factors which is called Outer OA. For this problem we can use $L_8$ as Inner OA and $L_4$ as Outer OA. The structure of the design matrix is given in Table 13.14. Depending on the number of control factors and the number of noise factors appropriate OAs are selected for design.

**TABLE 13.14** Inner/Outer OA design

$L_4$ OA (Outer Array)

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | *Z* | 1 | 2 | 2 | 1 |
| | | | | | | | *Y* | 1 | 2 | 1 | 2 |
| $L_8$ OA (Inner Array) | | | | | | | *X* | 1 | 1 | 2 | 2 |
| *Trial no.* | 1 *A* | 2 *B* | 3 *C* | 4 *D* | 5 *E* | 6 *F* | 7 *G* | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | * | * | * | * |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | * | * | * | * |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | * | * | * | * |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | * | * | * | * |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | * | * | * | * |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | * | * | * | * |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | * | * | * | * |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | * | * | * | * |

Note that the control factors are assigned to the inner array and noise factors are assigned to the outer array. There are 32 (8 × 4) separate test conditions (experiments). For example, observation $Y_1$ is obtained for Trial no. 1 by setting all control factors at first level and all noise factors at their first level and $Y_2$, $Y_3$ and $Y_4$ are obtained by changing the levels of noise factors alone. Thus, each observation is from one separate experiment. For each trial of inner OA, the S/N ratio is computed. Treating this S/N ratio as response, the data is analysed to arrive at the optimal levels for the control factors.

## ILLUSTRATION 13.2

**Inner/Outer OA: Smaller—the better type of quality characteristic**

Suppose an experiment was conducted using Inner/Outer OA and data were collected as given in Table 13.15. If the objective is to minimize the response, we use smaller—the better type of quality characteristic to compute S/N ratios [Eq. (13.1)]. That is

$$\eta = -10 \log\left[\frac{1}{n} \sum Y_i^2\right]$$

The S/N ratios obtained are given in Table 13.16. For example, for Trial 1, the S/N ratio is

$$\eta = -10 \log\left[\frac{1}{n} \sum Y_i^2\right]$$

$$= -10 \log\left[\frac{1}{4} \sum (10.3)^2 + (9.8)^2 + (10.8)^2 + (10.7)^2\right]$$

$$= -10 \log(108.315)$$

$$= -20.35$$

This S/N data can be analysed using either response graph method or ANOVA.

**TABLE 13.15** Data for Illustration 13.2

$L_4$ OA (Outer Array)

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | *Z* | 1 | 2 | 2 | 1 |
| | | | | | | | *Y* | 1 | 2 | 1 | 2 |
| $L_8$ OA (Inner Array) | | | | | | | *X* | 1 | 1 | 2 | 2 |
| *Trial* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | | | |
| *no.* | *A* | *B* | *C* | *D* | *E* | *F* | *G* | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10.3 | 9.8 | 10.8 | 10.7 |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 9.7 | 11.8 | 12.8 | 11. 4 |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 12.4 | 13.2 | 10.3 | 13.4 |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 9.4 | 9 | 8.6 | 9.4 |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 14.6 | 15.2 | 14.6 | 14.6 |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 8.5 | 9.6 | 6.2 | 8.5 |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 14.7 | 13 | 17.5 | 10.3 |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 9.4 | 12.3 | 10 | 8.6 |

**TABLE 13.16** S/N ratios for Illustration 13.2

| | | | | | | | Z | 1 | 2 | 2 | 1 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Y | 1 | 2 | 1 | 2 | |
| $L_8$ OA (Inner Array) | | | | | | | X | 1 | 1 | 2 | 2 | |
| *Trial no.* | 1 *A* | 2 *B* | 3 *C* | 4 *D* | 5 *E* | 6 *F* | 7 *G* | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ | $\eta$ |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10.3 | 9.8 | 10.8 | 10.7 | –20.35 |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 9.7 | 11.8 | 12.8 | 11.4 | –21.20 |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 12.4 | 13.2 | 10.3 | 13.4 | –21.85 |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 9.4 | 9.0 | 8.6 | 9.4 | –19.19 |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 14.6 | 15.2 | 14.6 | 14.6 | –23.38 |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 8.5 | 9.6 | 6.2 | 8.5 | –18.31 |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 14.7 | 13.0 | 17.5 | 10.3 | –23.00 |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 9.4 | 12.3 | 10.0 | 8.6 | –20.15 |

### *Data analysis using response graph method*

Now, we treat the S/N ratio as the response data and analyse it as discussed in Chapter 12. The level totals of the S/N ratios are given in Table 13.17.

**TABLE 13.17** Level totals of S/N ratios for Illustration 13.2

| *Factors* | *Level* 1 | *Level* 2 |
|---|---|---|
| *A* | –82.59 | –84.84 |
| *B* | –83.24 | –84.19 |
| *C* | –84.70 | –82.73 |
| *D* | –88.58 | –78.85 |
| *E* | –80.66 | –86.77 |
| *F* | –83.07 | –84.36 |
| *G* | –80.85 | –86.58 |

The grand total is –167.43 and the grand mean is –20.93 (–167.43/8). Table 13.18 gives the factor effects and their ranking.

**TABLE 13.18** Factor effects (average of S/N ratio) and ranks for Illustration 13.2

| *Factor* | *A* | *B* | *C* | *D* | *E* | *F* | *G* |
|---|---|---|---|---|---|---|---|
| Level 1 | –20.65 | –20.81 | –21.18 | –22.15 | –20.17 | –20.77 | –20.21 |
| Level 2 | –21.21 | –21.05 | –20.68 | –19.71 | –21.69 | –21.09 | –21.65 |
| Difference | 0.56 | 0.24 | 0.50 | 2.44 | 1.52 | 0.32 | 1.44 |
| Rank | 4 | 7 | 5 | 1 | 2 | 6 | 3 |

Figure 13.1 shows the response graph for Illustration 13.2.

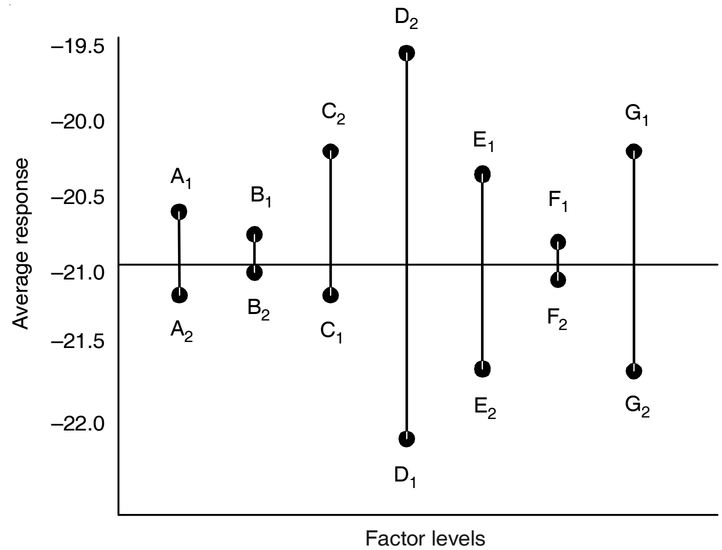



**FIGURE 13.1** Response graph for Illustration 13.2.

***Predicting optimum condition:*** From Table 13.18, it can be seen that factors $D$, $E$ and $G$ are significant (first three ranks). Hence the predicted optimum response in terms of S/N ratio ($\eta_{\text{opt}}$) is given by

$$\eta_{\text{opt}} = \bar{\eta} + (\bar{D}_2 - \bar{\eta}) + (\bar{E}_1 - \bar{\eta}) + (\bar{G}_1 - \bar{\eta}) \tag{13.5}$$

$$= \bar{D}_2 + \bar{E}_1 + \bar{G}_1 - 2 \times \bar{\eta}$$

$$= -19.71 - 20.17 - 20.21 - 2 \times (-20.93)$$

$$= -60.09 + 41.86$$

$$= -18.23$$

It is observed from Table 13.16 that the maximum S/N ratio is –18.31 and it corresponds to the experimental Trial no. 6. The optimum condition found is $D_2\, E_1\, G_1$. And the experimental Trial no. 6 includes this condition with a predicted yield of –18.23. From this we can conclude that the optimum condition obtained is satisfactory. However, the result has to be verified through a confirmation experiment.

### Data analysis using ANOVA

The sum of squares is computed using the level totals (Table 13.17).

$$\text{Correction factor (CF)} = \frac{(-167.43)^2}{8} = 3504.10$$

$$SS_A = \frac{(-82.59)^2 + (-84.84)^2}{4} - \text{CF}$$

$$= 3504.73 - 3504.10$$

$$= 0.63$$

Similarly, all factor sum of squares are obtained.

$SS_B = 0.11$, $SS_C = 0.49$, $SS_D = 11.83$, $SS_E = 4.67$, $SS_F = 0.21$, $SS_G = 4.10$

$SS_{\text{Total}} = 22.04$

These are summarised in Table 13.19. Since we have only one replicate, experimental error (pure error) will be zero. And we have to use pooled error for testing the effects which is shown in Table 13.19. At 5% level of significance, it is observed that factors *D*, *E* and *G* are significant. And these three factors together account for about 93% of total variation. This result is same as that obtained in the response graph method. The determination of optimal levels for these factors and $\eta_{\text{Opt}}$ is similar to that of response graph method.

**TABLE 13.19** ANOVA for Illustration 13.2

| *Source of variation* | *Sum of squares* | *Degree of freedom* | *Mean square* | $F_0$ | *C(%)* |
|---|---|---|---|---|---|
| *A** | 0.63 | 1 | – | – | 2.86 |
| *B** | 0.11 | 1 | – | – | 0.50 |
| *C** | 0.49 | 1 | – | – | 2.22 |
| *D* | 11.83 | 1 | 11.83 | 32.86 | 53.68 |
| *E* | 4.67 | 1 | 4.67 | 12.97 | 21.19 |
| *F** | 0.21 | 1 | – | – | 0.95 |
| *G* | 4.10 | 1 | 4.10 | 11.39 | 18.60 |
| Pooled error | 1.44 | 4 | 0.36 | 6.53 | |
| Total | 22.04 | 7 | | | 100.00 |

*Pooled into error

## ILLUSTRATION 13.3

**Inner/Outer OA: Larger—the better type of quality characteristic**

In the machining of a metal component, the following control and noise factors were identified as affecting the surface hardness.

*Control factors:* Speed (*A*), Feed (*B*), Depth of cut (*C*) and Tool angle (*D*)

*Noise factors:* Tensile strength of material (*X*), Cutting time (*Y*) and Operator skill (*Z*)

Control factors were studied each at three levels and noise factors each at two levels. $L_9$ OA was used as inner array and $L_4$ as outer array. The control factors were assigned to the inner array and the noise factors to the outer array. The objective of this study is to maximize the surface hardness. The data collected (hardness measurements) along with the design is given in Table 13.20.

**TABLE 13.20** Design and data for Illustration 13.3

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | *Z* | 1 | 1 | 2 | 2 | |
| | | | | *Y* | 1 | 2 | 1 | 2 | |
| | | | | *X* | 1 | 2 | 2 | 1 | |
| *Trial no.* | 1 | 2 | 3 | 4 | | | | | S/N ratio |
| | *A* | *B* | *C* | *D* | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ | ($\eta$) |
| 1 | 1 | 1 | 1 | 1 | 273 | 272 | 285 | 284 | 48.89 |
| 2 | 1 | 2 | 2 | 2 | 253 | 254 | 270 | 267 | 48.31 |
| 3 | 1 | 3 | 3 | 3 | 229 | 231 | 260 | 260 | 47.73 |
| 4 | 2 | 1 | 2 | 3 | 223 | 221 | 240 | 243 | 47.27 |
| 5 | 2 | 2 | 3 | 1 | 260 | 260 | 290 | 292 | 48.76 |
| 6 | 2 | 3 | 1 | 2 | 219 | 216 | 245 | 245 | 47.23 |
| 7 | 3 | 1 | 3 | 2 | 256 | 257 | 290 | 290 | 48.68 |
| 8 | 3 | 2 | 1 | 3 | 189 | 190 | 210 | 215 | 46.02 |
| 9 | 3 | 3 | 2 | 1 | 270 | 267 | 290 | 290 | 48.90 |

Total = 431.79
Mean = 47.98

***Data analysis:*** The data of Illustration 13.3 is analysed by ANOVA. For each experiment (trial no.), there are four observations ($Y_1$, $Y_2$, $Y_3$ and $Y_4$). Using this data, for each trial, the S/N ratio is computed. Since the objective is to maximize the hardness, the objective function is larger the better type of quality characteristic for which the S/N ratio is given by Eq. (13.3). For the first trial, the data are 273, 272, 285 and 284. Substituting the data in Eq. (13.3), we get

$$\eta_1 = -10 \log \left( \frac{1}{n} \sum \frac{1}{Y_i^2} \right) = -10 \log \left[ \frac{1}{4} \left( \frac{1}{(273)^2} + \frac{1}{(272)^2} + \frac{1}{(285)^2} + \frac{1}{(284)^2} \right) \right]$$
$$= 48.89$$

Similarly, for all trials S/N ratios are computed and given in Table 13.20. This S/N ratio data is analysed through ANOVA.

***Computation of sum of squares and ANOVA:*** The level totals of S/N ratios are given in Table 13.21.

**TABLE 13.21** Level totals of S/N ratios for Illustration 13.3

| *Factor* | *Level* 1 | *Level* 2 | *Level* 3 |
|---|---|---|---|
| *A* | 144.93 | 143.26 | 143.60 |
| *B* | 144.84 | 143.09 | 143.86 |
| *C* | 142.14 | 144.48 | 145.17 |
| *D* | 146.55 | 144.22 | 141.02 |

$$\text{Correction factor (CF)} = \frac{(431.79)^2}{9} = 20715.84$$

$$SS_A = \frac{A_1^2 + A_2^2 + A_3^2}{3} - \text{CF} \tag{13.6}$$

$$= \frac{(144.93)^2 + (143.26)^2 + (143.60)^2}{3} - 20715.84 = 0.52$$

Similarly, $SS_B = 0.52,\ SS_C = 1.68,\ SS_D = 5.14$

As we have only one replicate, experimental error (pure error) will be zero. And we have to use pooled error for testing the effects. The ANOVA for Illustration is given in Table 13.22. At 5% level of significance only the tool angle has significant influence on the surface hardness. And the depth of cut also has considerable influence (21% contribution) on the hardness.

**TABLE 13.22** ANOVA for Illustration 13.3

| *Source of variation* | *Sum of squares* | *Degree of freedom* | *Mean squares* | $F_0$ | *C*(%) |
|---|---|---|---|---|---|
| Speed (*A*)* | 0.52 | 2 | – | – | 6.62 |
| Feed (*B*)* | 0.52 | 2 | – | – | 6.62 |
| Depth of cut (*C*) | 1.68 | 2 | 0.84 | 3.23 | 21.37 |
| Tool angle (*D*) | 5.14 | 2 | 2.57 | 9.88 | 65.39 |
| Pooled error | 1.04 | 4 | 0.26 | | |
| Total | 7.86 | 8 | | | |

*Pooled into error

***Optimal levels for the control factors:*** The average response in terms of S/N ratio of each level of all factors is given in Table 13.23.

**TABLE 13.23** Average response (S/N ratio) for Illustration 13.3

| *Factors* | *Level* 1 | *Level* 2 | *Level* 3 |
|---|---|---|---|
| *A* | **48.31** | 47.75 | 47.87 |
| *B* | **48.28** | 47.70 | 47.95 |
| *C* | 47.38 | 48.16 | **48.39** |
| *D* | **48.85** | 48.07 | 47.01 |

The optimal levels for all the factors are always selected based on the maximum average S/N ratio only irrespective of the objective (maximization or minimization) of the problem. Accordingly, the best levels for the factors are $A_1\ B_1\ C_3\ D_1$.

***Predicting the optimum response:*** Since the contribution of $A$ and $B$ is negligible, we can consider only factors $C$ and $D$ for predicting the optimum response.

$$\eta_{\text{opt}} = \bar{\eta} + (\bar{C}_3 - \bar{\eta}) + (\bar{D}_1 - \bar{\eta}) \tag{13.7}$$

$$= 47.98 + (48.39 - 47.98) + (48.85 - 47.98)$$

$$= 49.26$$

## 13.7 RELATION BETWEEN S/N RATIO AND QUALITY LOSS

When experiments are conducted on the existing process/product, it is desirable to assess the benefit obtained by comparing the performance at the optimum condition and the existing condition.

One of the measures of performance recommended in robust design is the gain in loss per part ($g$).

It is the difference between loss per part at the optimum condition and the existing condition.

$$g = L_{\text{ext}} - L_{\text{opt}}$$

$$= K(\text{MSD}_{\text{ext}} - \text{MSD}_{\text{opt}}) \tag{13.8}$$

It can also be expressed in terms of S/N ratio ($\eta$) = –10 log (MSD).

Let $\eta_e$ = S/N ratio at the existing condition
$\eta_o$ = S/N ratio at the optimum condition
$M_e$ = MSD at the existing condition
$M_o$ = MSD at the optimum condition

Therefore,

$$\eta_o = -10 \log M_o$$

$$M_o = 10^{\frac{-\eta_o}{10}} \text{ and } M_e = 10^{\frac{-\eta_e}{10}}$$

If $R_L$ is the proportion of loss reduction, then

$$R_L = \frac{M_o}{M_e} = 10^{-\left(\frac{\eta_o + \eta_e}{10}\right)} = 10^{\frac{-X}{10}}$$

where $X = \eta_o - \eta_e$ .

It is the gain in signal ratio.

We can also express $R_L$ in another way

Suppose $R_L = 0.5^{X/K}$

Then, $10^{-X/10} = 0.5^{X/K}$

That is, $10^{-1/10} = 0.5^{1/K}$

Taking logarithm on both sides and simplifying, we obtain

$$K = -10 \log 0.5 = 3.01, \text{ approximately } 3.0$$

Therefore, reduction in loss ($R_L$) = $10^{-X/10} = 0.5^{X/3} = 0.5^{\frac{\eta_o - \eta_e}{3}}$ (13.9)

$R_L$ is a factor by which the $MSD_{ext}$ has been reduced.

Since
$$R_L = \frac{MSD_{opt}}{MSD_{ext}}$$

$$MSD_{opt} = R_L(MSD_{ext})$$

and
$$g = K(MSD_{ext} - MSD_{opt})$$
$$= K(MSD_{ext} - R_L * MSD_{ext})$$
$$= K.\ MSD_{ext}\ (1 - R_L)$$
$$= K.\ MSD_{ext}\left(1 - 0.5^{\frac{\eta_o - \eta_e}{3}}\right) \tag{13.10}$$

Note that the quantity $K.\ MSD_{ext}$ is the loss at the original (existing) condition. Suppose the value of the quantity in the parenthesis of Eq. (13.10) is 0.40.

That is,
$$g = K.\ MSD_{ext.}(0.40)$$

This indicates that a savings of 40% of original loss to the society is achieved.

## ILLUSTRATION 13.4

**Estimation of Quality Loss**

In Illustration 13.2, we have obtained $\eta_{opt}$ as –18.23 [Eq. (13.5)]. For the same problem in Illustration 13.2, suppose the user is using $D_1$, $E_1$, and $G_1$ at present. For this existing condition we can compute the S/N ratio ($\eta_{ext}$). By comparing this with $\eta_{opt}$, we can compute gain in S/N ratio. Using data from Table 13.18,

$$\eta_{ext} = \bar{\eta} + (\bar{D}_1 - \bar{\eta}) + (\bar{E}_1 - \bar{\eta}) + (\bar{G}_1 - \bar{\eta}) \tag{13.11}$$
$$= \bar{D}_1 + \bar{E}_1 + \bar{G}_1 - 2 \times \bar{\eta}$$
$$= -22.15 - 20.17 - 20.21 - 2 \times (-20.93)$$
$$= -62.53 + 41.86$$
$$= -20.67$$

Therefore, the reduction in loss ($R_L$) to society is computed from Eq. (13.10).

$$g = K\ MSD_{ext}\left[1 - 0.5^{\left(\frac{\eta_{opt} - \eta_{ext}}{3}\right)}\right]$$

$$g = K\ MSD_{ext}\left[1 - 0.5^{\left(\frac{-18.23 + 20.67}{3}\right)}\right]$$

or
$$g = K\ \text{MSD}_{\text{ext}}\ (1 - 0.5^{0.81})$$
$$= K\ \text{MSD}_{\text{ext}}\ (1 - 0.57)$$
$$= K\ \text{MSD}_{\text{ext}} \times 0.43$$

This indicates that there is a saving of 43% of the original loss to society.

## PROBLEMS

**13.1** An experiment was conducted with seven main factors ($A$, $B$, $C$, $D$, $E$, $F$, and $G$) using $L_8$ OA and the following data was collected (Table 13.24). Assuming larger—the better type quality characteristic, compute S/N ratios and identify the optimal levels for the factors.

**TABLE 13.24** Data for the Problem 13.1

| *Trial* | *Factors/Columns* | | | | | | | *Response* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *no.* | *A* 1 | *B* 2 | *C* 3 | *D* 4 | *E* 5 | *F* 6 | *G* 7 | $R_1$ | $R_2$ | $R_3$ |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 11 | 4 | 11 |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 4 | 4 | 4 |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 4 | 1 | 14 |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 4 | 0 | 8 |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 9 | 8 | 4 |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 4 | 1 | 1 |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 4 | 4 |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 14 | 4 | 8 |

**13.2** Suppose data is available from an Inner/Outer OA experiment (Table 13.25).
(a) Assuming smaller—the better type of quality characteristic, compute S/N ratios.
(b) Analyse S/N data using response graph method and determine significant effects.
(c) Compute S/N ratio at the optimum condition ($\eta_{\text{opt}}$).
(d) Assuming that all the significant factors are currently at their first level, compute S/N ratio for the existing condition ($\eta_{\text{ext}}$).
(e) Relate the gain in loss to society to the gain in signal to noise ratio.

**TABLE 13.25** Data for Problem 13.2

| | | | | | | | | $L_4$ OA (Outer Array) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | *Z* | 1 | 2 | 2 | 1 |
| | | | | | | | *Y* | 1 | 2 | 1 | 2 |
| $L_8$ OA (Inner Array) | | | | | | | *X* | 1 | 1 | 2 | 2 |
| *Trial* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | | | |
| *no.* | *A* | *B* | *C* | *D* | *E* | *F* | *G* | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 39 | 38 | 35 | 36 |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 52 | 53 | 55 | 50 |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 31 | 36 | 34 | 35 |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 45 | 42 | 43 | 45 |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 35 | 32 | 33 | 37 |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 23 | 22 | 21 | 25 |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 18 | 14 | 20 | 21 |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 26 | 23 | 25 | 22 |

**13.3** Consider the data in Problem 13.2.

(a) Assume larger—the better type of quality characteristic and compute S/N ratios.
(b) Compute sum of squares of all factors.
(c) Perform ANOVA and determine the significant effects.

CHAPTER

# 14 Multi-level Factor Designs

## 14.1 INTRODUCTION

Usually, we make use of the standard Orthogonal Arrays (OAs) for designing an experiment. When these standard arrays are not suitable for studying the problem under consideration, we may have to modify the standard orthogonal arrays. Suppose we have a problem where one factor has to be studied at four-level and others at two-level. There is no standard OA which can accommodate one four-level factor along with two-level factors or a three-level factor with two-level factors. To deal with such situations we have to modify a two-level standard OA to accommodate three-level or four-level factors. These modified OAs when used for designing an experiment are termed multi-level factor designs. When we deal with discrete factors and continuous variables, we encounter with multi-level factor designs. In this chapter some of the methods used for modification of standard OAs and their application are discussed.

## 14.2 METHODS FOR MULTI-LEVEL FACTOR DESIGNS

The methods which are often employed for accommodating multiple levels are as follows:

**1.** ***Merging of columns:*** In this method two or more columns of a standard OA are merged to create a multi-level column. For example, a four-level factor can be fit into a two-level OA.

**2.** ***Dummy treatment:*** In this method, one of the levels of a higher level factor is treated as dummy when it is fit into a lesser number level OA. For example, accommodating a three-level factor into a two-level OA.

**3.** ***Combination method:*** This method can be used to fit two-level factors into a three-level OA.

**4.** ***Idle column method:*** This method can be used to accommodate many three-level factors into a two-level OA. The first two methods do not result in loss of orthogonality of the entire experiment. Methods 3 and 4 cause a loss of orthoganality among the factors studied by these methods and hence loss of an accurate estimate of the independent factorial effects. The advantage of method 3 and 4 is a smaller experiment.

### 14.2.1 Merging Columns

Merging columns method is more useful to convert a two-level OA to accommodate some four-level columns. Suppose we want to introduce a four-level column into a two-level OA. The four-level factors will have 3 degrees of freedom (df) and a two-level column has 1 degree of freedom. So, we have to merge (combine) three two-level columns to create one four-level column. It is recommended to merge three mutually interactive columns, such as 1, 2 and 3 or 2, 6 and 4 of $L_8$ OA. The merging of mutually interactive columns minimizes the confounding of interactions. For creating the four-level column, the following procedure is followed.

In a two-level OA, any two columns will contain the pairs (1, 1), (1, 2), (2, 1) and (2, 2). The four-level column is created by assigning level 1 to pair (1, 1) level 2 to (1, 2), level 3 to (2, 1) and level 4 to (2, 2). Table 14.1 gives the merging of columns (1, 2 and 3) of $L_8$ OA to create a four-level column.

**TABLE 14.1** Merging columns 1, 2 and 3 of $L_8$ OA

| *Trial no.* | *Columns* 1 | 2 | 3 | *Four-level column* |
|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 |
| 3 | 1 | 2 | 2 | 2 |
| 4 | 1 | 2 | 2 | 2 |
| 5 | 2 | 1 | 2 | 3 |
| 6 | 2 | 1 | 2 | 3 |
| 7 | 2 | 2 | 1 | 4 |
| 8 | 2 | 2 | 1 | 4 |

Suppose factor $A$ is assigned to the four-level column. The computation of sum of squares for the four-level column is

$$SS_A = \frac{A_1^2 + A_2^2 + A_3^2 + A_4^2}{n_A} - \text{CF} \tag{14.1}$$

where $n_A$ is the number of observations in the total $A_i$.

The $L_8$ OA modified with a four-level columns is given in Table 14.2.

**TABLE 14.2** $L_8$ OA modified for a four-level factor

| *Trial no.* | *Columns* 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 2 | 2 | 2 | 2 |
| 3 | 2 | 1 | 1 | 2 | 2 |
| 4 | 2 | 2 | 2 | 1 | 1 |
| 5 | 3 | 1 | 2 | 1 | 2 |
| 6 | 3 | 2 | 1 | 2 | 1 |
| 7 | 4 | 1 | 2 | 2 | 1 |
| 8 | 4 | 2 | 1 | 1 | 2 |

Table 14.2 has been obtained by merging columns 1, 2 and 3 of standard $L_8$ OA. Other combinations of the four-level factor can be obtained by selecting any other set of three mutually interactive columns from the triangular table of $L_8$ OA (Appendix B).

***Interaction between the four-level and two-level factor***

Suppose we assign factor $A$ to the four-level column obtained by merging columns 1, 2 and 3 of standard $L_8$ OA and factor $B$ to the column 4 of the standard $L_8$ OA. Figure 14.1 shows the linear graph for the $AB$ interaction. The Interaction sum of squares $SS_{AB}$ is given by the sum of squares of columns 5, 6 and 7 (Figure 14.1). That is,

$$SS_{AB} = SS_5 + SS_6 + SS_7 \tag{14.2}$$

The interaction degrees of freedom = 3 × 1 = 3 (equal to the sum of df of columns 5, 6 and 7), like this, we can also modify the standard $L_{16}$ OA to accommodate more than one four-level factor.

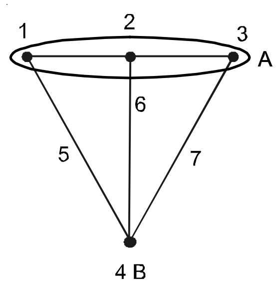



FIGURE 14.1 Interaction between a four-level factor and a two-level factor.

### 14.2.2 Dummy Treatment

Any factor has to be studied with a minimum of two levels in order to obtain the factor effect. So, to make one of the levels of a factor as dummy, the OA should have at least three levels. Accordingly, dummy level designs can be used with columns that have three or more levels. For example, we can assign a two-level factor to a three-level column of $L_9$ OA and treat one of the levels as dummy. Similarly, a three-level factor can be assigned to a four-level column and treat one level as dummy.

***Conversion of $L_9$ OA to accommodate one two-level factor***

Suppose, we want to study four factors, one with two levels and the other three factors each at three levels. The appropriate OA here is $L_9$. When we want to assign a two-level factor ($A$) to the first column of the standard $L_9$ OA, it is modified with a dummy treatment as given in Table 14.3.

Note that level 3 is dummy treated as given in Table 14.3. We can use any one level as dummy. The factor $A$ has 1 degree of freedom, whereas the column has 2 degrees of freedom. So, this extra one degree of freedom has to be combined with the error. The sum of squares of factor $A$ is given by

$$SS_A = \frac{(A_1 + A_1')^2}{n_{A_1} + n_{A_1'}} + \frac{A_2^2}{n_{A_2}} - \text{CF} \tag{14.3}$$

The error sum of squares associated with the dummy column is given by

$$SS_{eA} = \frac{(A_1 - A_1')^2}{n_{A_1} + n_{A_1}'} \tag{14.4}$$

This error ($SS_{eA}$) has to be added with the experimental error (pure error) along with its degree of freedom (1 df).

**TABLE 14.3** Dummy level design of $L_9$ OA

| *Trial no.* | 1(*A*) | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 2 | 2 | 2 |
| 3 | 1 | 3 | 3 | 3 |
| 4 | 2 | 1 | 3 | 3 |
| 5 | 2 | 2 | 2 | 1 |
| 6 | 2 | 3 | 1 | 2 |
| 7 | 1′ | 1 | 3 | 2 |
| 8 | 1′ | 2 | 1 | 3 |
| 9 | 1′ | 3 | 2 | 1 |

### *Conversion of a two-level OA to accommodate a three-level factor*

It involves the following two-step procedure

*Step* 1: Obtain a four-level factor from a two-level OA.
*Step* 2: Using dummy treatment, obtain three-level factor from the possible four levels.

Suppose we want to create a three-level column in $L_8$ OA. First, we create a four-level column using the method of merging columns. Then using dummy treatment method, the four-level column is converted into a three-level factor column. This is given in Table 14.4.

**TABLE 14.4** $L_8$ OA Dummy treatment for factor *A* (three-level)

| *Four-level* 1, 2, 3 | *Alternate three-level options from the four-level* | | |
|---|---|---|---|
| 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | 1 |
| 2 | 2 | 2 | 2 |
| 2 | 2 | 2 | 2 |
| 3 | 3 | 3 | 3 |
| 3 | 3 | 3 | 3 |
| 4 | 1′ | 2′ | 3′ |
| 4 | 1′ | 2′ | 3′ |

Suppose factor *A* with three levels is assigned to the dummy column. This column has 3 degrees of freedom whereas the factor *A* has 2 degrees of freedom. So, the extra one degree of freedom is to be combined with the experimental error (pure error). The sum of squares is computed as follows:

$$SS_A = \frac{(A_1 + A_1')^2}{n_{A_1} + n'_{A_1}} + \frac{A_2^2}{n_{A_2}} + \frac{A_3^2}{n_{A_3}} - \text{CF} \tag{14.5}$$

$$SS_{e_A} = \frac{(A_1 - A_1')^2}{n_{A_1} + n'_{A_1}} \tag{14.6}$$

***Data analysis from dummy-level design:*** Suppose, we have collected the data using the following dummy-level design (Table 14.5).

**TABLE 14.5** Data for dummy-level design

| *Experiment no.* | *A* 1 | *B* 2 | *C* 3 | *D* 4 | *Response* $R_1$ | $R_2$ |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 15 | 13 |
| 2 | 1 | 2 | 2 | 2 | 4 | 3 |
| 3 | 1 | 3 | 3 | 3 | 0 | 1 |
| 4 | 2 | 1 | 2 | 3 | 6 | 7 |
| 5 | 2 | 2 | 3 | 1 | 6 | 5 |
| 6 | 2 | 3 | 1 | 2 | 12 | 14 |
| 7 | 1′ | 1 | 3 | 2 | 6 | 5 |
| 8 | 1′ | 2 | 1 | 3 | 9 | 10 |
| 9 | 1′ | 3 | 2 | 1 | 0 | 2 |

The analysis of the data using ANOVA is explained below. The response totals for Illustration 14.1 is given in Table 14.6.

**TABLE 14.6** Response totals for the dummy-level design

| *Level* | *Factors* *A* | *B* | *C* | *D* |
|---|---|---|---|---|
| 1 | 36 | 52 | 73 | 41 |
| 2 | 50 | 37 | 22 | 44 |
| 3 | – | 29 | 23 | 33 |
| 1′ | 32 | – | – | – |

Grand total ($T$) = 118, $N$ = 18

$$\text{CF} = \frac{T^2}{N} = \frac{118^2}{18} = 773.56$$

$$SS_A = \frac{(A_1 + A_1')^2}{n_{A_1} + n_{A_1'}} + \frac{A_2^2}{n_{A_2}} - \text{CF}$$

$$= \frac{68^2}{6+6} + \frac{50^2}{6} - \text{CF}$$

$$= 385.33 + 416.67 - 773.56$$

$$= 28.44$$

$$SS_B = \frac{B_1^2 + B_2^2 + B_3^2}{n_B} - \text{CF}$$

$$= \frac{52^2 + 37^2 + 29^2}{6} - 773.56$$

$$= 45.44$$

Similarly, $SS_C = 283.44$ and $SS_D = 10.78$

$$SS_{\text{Total}} = 15^2 + 4^2 + \cdots + 10^2 + 2^2 - \text{CF}$$

$$= 1152.00 - 773.56$$

$$= 378.44$$

$$SS_{eA} = \frac{(A_1 - A_1')^2}{n_{A_1} + n'_{A_1}} = \frac{(36 - 32)^2}{6+6} = \frac{16}{12} = 1.33$$

$$SS_e = SS_{\text{Total}} - SS_A - SS_B - SS_C - SS_D - SS_{e_A}$$

$$= 378.44 - 28.44 - 45.44 - 283.44 - 10.78 - 1.33 = 9.01$$

The computations are summarized in the ANOVA Table 14.7.

**TABLE 14.7** ANOVA for the dummy-level design

| *Source* | *Sum of squares* | *Degrees of freedom* | *Mean squares* | $F_0$ | *Significance* |
|---|---|---|---|---|---|
| $A$ | 28.44 | 1 | 28.44 | 16.17 | Significant |
| $B$ | 45.44 | 2 | 22.72 | 12.92 | Significant |
| $C$ | 283.44 | 2 | 141.72 | 80.56 | Significant |
| $D^*$ | 10.78 | 2 | _ | _ | |
| $e_A^*$ | 1.33 | 1 | _ | _ | |
| Error | 9.01 | 9 | _ | _ | |
| $e_{\text{pooled}}$ | 21.12 | 12 | 1.76 | _ | |
| Total | 378.44 | 17 | | | |

* Pooled into error; $F_{5\%,1,12} = 4.75$, $F_{5\%,2,12} = 3.89$

### 14.2.3 Combination Method

Column merging and dummy treatment methods are widely used. Only under certain circumstances combination method is used. Suppose we want to study a mixture of two-level and three-level factors. We can use a two-level OA or a three-level OA with dummy treatment method depending on the number of two-level or three-level factors. If the number of factors is at the limit of the number of columns available in an OA and the next larger OA is not economical to use and we do not want to eliminate factors to fit the OA, we use the combination method. In this method a pair of two-level factors is treated as a three-level factor. The interaction between these two factors cannot be studied. How to combine two two-level factors $A$ and $B$ to study their effects is explained as follows:

The four possible combinations of two-level factors $A$ and $B$ are

$$A_1B_1,\ A_1B_2,\ A_2B_1 \text{ and } A_2B_2$$

Suppose, we select the test conditions $A_1B_1$, $A_2B_2$ and one of the remaining two, say, $A_2B_1$. Now by comparing $A_1B_1$ with $A_2B_1$, the effect of $A$ can be estimated (factor $B$ is kept constant). Similarly, by comparing $A_2B_2$ with $A_2B_1$, the effect of factor $B$ can be estimated (factor $A$ is kept constant). However, interaction between $A$ and $B$ cannot be studied. These three test conditions ($A_1B_1$, $A_2B_2$ and $A_2B_1$) can be assigned to one column in a three-level OA as a combined factor $AB$ (Table 14.8). Because one test condition $A_1B_2$ is not tested, the orthogonality between $A$ and $B$ is lost.

**TABLE 14.8** Combined factors assignment

| *Trial no.* | *Factors/columns* | | | |
|---|---|---|---|---|
| | *AB* | *C* | *D* | *E* |
| | 1 | 2 | 3 | 4 |
| 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 2 | 2 | 2 |
| 3 | 1 | 3 | 3 | 3 |
| 4 | 2 | 1 | 2 | 3 |
| 5 | 2 | 2 | 3 | 1 |
| 6 | 2 | 3 | 1 | 2 |
| 7 | 3 | 1 | 3 | 2 |
| 8 | 3 | 2 | 1 | 3 |
| 9 | 3 | 3 | 2 | 1 |

The three combinations $A_1B_1$, $A_2B_1$ and $A_2B_2$ selected are assigned to the column 1 of $L_9$ OA as follows. $AB_1$ (Table 14.8) is assigned to $A_1B_1$, $AB_2$ is assigned to $A_2B_1$ and $AB_3$ is assigned to $A_2B_2$ (Table 14.9). Table 14.9 also gives the assignment of three-level factors $C$, $D$ and $E$.

**TABLE 14.9** Combined factor layout design

| *Trial no.* | *Factors/Columns* | | | | | *Response (Y)* |
|---|---|---|---|---|---|---|
| | *AB* 1 | | *C* 2 | *D* 3 | *E* 4 | |
| 1 | 1 | 1 | 1 | 1 | 1 | 4 |
| 2 | 1 | 1 | 2 | 2 | 2 | 7 |
| 3 | 1 | 1 | 3 | 3 | 3 | 10 |
| 4 | 2 | 1 | 1 | 2 | 3 | 5 |
| 5 | 2 | 1 | 2 | 3 | 1 | 9 |
| 6 | 2 | 1 | 3 | 1 | 2 | 1 |
| 7 | 2 | 2 | 1 | 3 | 2 | 14 |
| 8 | 2 | 2 | 2 | 1 | 3 | 5 |
| 9 | 2 | 2 | 3 | 2 | 1 | 11 |

Total = 66

***Data analysis from combined factor design:*** Suppose, the last column of Table 14.9 gives the experimental results. The data are analysed as follows. The response totals for all the factors is given in Table 14.10.

**TABLE 14.10** Response totals for the combined design

| *Level* | *Factors (effects)* | | | | | |
|---|---|---|---|---|---|---|
| | *AB* | *A* ($B_1$ constant) | *B* ($A_2$ constant) | *C* | *D* | *E* |
| 1 | 21 | 21($A_1B_1$) | 15($A_2B_1$) | 23 | 10 | 24 |
| 2 | 15 | 15($A_2B_1$) | 30($A_2B_2$) | 21 | 23 | 22 |
| 3 | 30 | — | — | 22 | 33 | 20 |

Note that the totals of $A$ and $B$ are obtained such that the individual effects of $A$ and $B$ are possible. That is,

$A_1$ = 21 (combination of $A_1B_1$) and
$A_2$ = 15 (combination of $A_2B_1$)

Since $B_1$ is kept constant in these two combinations, effect of $A$ can be obtained. Similarly, total for $B_1$ and $B_2$ are obtained from $A_2B_1$ and $A_2B_2$.

***Computation of sum of squares:***

$$\text{CF} = \frac{T^2}{N} = \frac{(66)^2}{9} = 484.00$$

$$SS_{\text{Total}} = (4)^2 + (7)^2 + \ldots + (11)^2 - \text{CF}$$
$$= 614 - 484 = 130.00$$

$$SS_{AB} = \frac{(AB_1)^2 + (AB_2)^2 + (AB_3)^2}{n_{AB}} - \text{CF} \quad (AB \text{ totals are arrived from Table 14.10})$$

$$= \frac{(21)^2 + (15)^2 + (30)^2}{3} - 484.00$$

$$= 522 - 484 = 38.00$$

Similarly, $SS_C$, $SS_D$ and $SS_E$ are computed

$$SS_C = \frac{(23)^2 + (21)^2 + (22)^2}{3} - 484 = 484.66 - 484 = 0.66$$

$$SS_D = \frac{(10)^2 + (23)^2 + (33)^2}{3} - 484 = 572.67 - 484 = 88.67$$

$$SS_E = \frac{(24)^2 + (22)^2 + (20)^2}{3} - 484 = 486.67 - 484 = 2.67$$

$$SS_e = SS_{\text{Total}} - SS_{AB} - SS_C - SS_D - SS_E$$
$$= 130 - (38 + 0.66 + 88.67 + 2.67)$$
$$= 130 - 130 = 0.00$$

Since only one replication is taken the experimental error ($SS_e$) is obviously zero. In order to test the individual effects of $A$ and $B$, we need to compute $SS_A$ and $SS_B$.

$$SS_A = \frac{(A_1)^2 + (A_2)^2}{n_A} - \frac{(A_1 + A_2)^2}{n_{A_1} + n_{A_2}}$$

$$= \frac{(21)^2 + (15)^2}{3} - \frac{(36)^2}{6}$$

$$= \frac{666}{3} - \frac{1296}{6} = 222 - 216 = 6.00$$

$$SS_B = \frac{(B_1)^2 + (B_2)^2}{n_B} - \frac{(B_1 + B_2)^2}{n_{B_1} + n_{B_2}}$$

$$= \frac{(15)^2 + (30)^2}{3} - \frac{(45)^2}{6} = 375 - 337.5 = 37.5$$

For orthogonality between $A$ and $B$, we must have

$$SS_{AB} = SS_A + SS_B$$

But here we have, $$SS_{AB} \neq SS_A + SS_B$$

$$38 \neq 6 + 37.5$$

That is, $A$ and $B$ are not orthogonal. And it is so in the combination design. To take care of this, a correction to the sum of squares has to be made in the ANOVA. Suppose the difference between $SS_{AB}$ and $(SS_A + SS_B)$ is $\Delta_{AB}$.

$$\Delta_{AB} = SS_{AB} - (SS_A + SS_B)$$
$$= 38 - (6 + 37.5) = -5.5$$

Note that there is no degree of freedom associated with $\Delta_{AB}$ because $AB$ has 2 degrees of freedom and $A$ and $B$ has 1 degree of freedom each. The analysis of variance is given in Table 14.11.

**TABLE 14.11** ANOVA for the combined factor design

| *Source* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ | *Significance* |
|---|---|---|---|---|---|
| $AB$ | 38.00 | 2 | 19 | 24.6 | Significant |
| $A^*$ | 6.00 | 1 | — | | |
| $B$ | 37.5 | 1 | 37.5 | 48.7 | Significant |
| $\Delta_{AB}^*$ | –5.5 | 0 | — | | |
| $C^*$ | 0.66 | 2 | — | | |
| $D$ | 88.67 | 2 | 44.33 | 57.6 | Significant |
| $E^*$ | 2.67 | 2 | — | | |
| Pooled | 3.83 | 5 | 0.77 | | |
| Total | 130.00 | 8 | | | |

* Pooled into error; $F_{5\%,1,5} = 6.61$, $F_{5\%,2,5} = 5.79$

### 14.2.4 Idle Column Method

The idle column method is used to assign three-level factors in a two-level OA. A three-level factor can be assigned in a two-level OA by creating a four-level column and making 1 level as dummy. This design is not efficient since we waste 1 degree of freedom in the error due to the dummy level. But the idle column design is more efficient even though one column is left idle (no assignment). For two or more three-level factors, idle column method reduces the size of the experiment but sacrifies the orthogonality of the three-level factors.

The idle column should always be the first column of the two-level OA. The levels of this column are the basis for assigning the three-level factors. If a three-level factor say, $A$ is assigned to column 2 of $L_8$ OA, the interaction column 3 is dropped. Similarly, if another factor $B$ with three-levels is assigned to column 4, the interaction column 5 is eliminated. The idle column is the common column in the mutually inter active groups where three-level factors are assigned.

As already mentioned, always the first column of the two-level OA is the idle column. Its levels serve as the basis for the assignment of the three-level factors.

For example, against level 1 of the idle column we assign level 1 and 2 of the three-level factor and against level 2 of the idle column we assign levels 2 and 3 of the three-level factor. Table 14.12 gives the idle column assignment of one three-level factor ($A$).

**TABLE 14.12** Idle column assignment in $L_8$ OA

| *Trial no.* | *Factors/Columns* | | |
|---|---|---|---|
| | *Idle* | *A* | |
| | 1 | 2 | 3 |
| 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 |
| 3 | 1 | 2 | 2 |
| 4 | 1 | 2 | 2 |
| 5 | 2 | 2 | 2 |
| 6 | 2 | 2 | 2 |
| 7 | 2 | 3 | 1 |
| 8 | 2 | 3 | 1 |

Suppose we have two three-level factors ($A$ and $B$) and two two-level factors ($C$ and $D$). To study these factors, we require 6 degrees of freedom. Hence, we can use $L_8$ OA for assignment. Table 14.13 gives the assignment of these factors and data for illustration.

**TABLE 14.13** Idle column factor assignment

| *Trial no.* | *Factors/Columns* | | | | | *Response* |
|---|---|---|---|---|---|---|
| | *Idle* | *A* | *B* | *C* | *D* | |
| | 1 | 2 | 3 | 4 | 5 | |
| 1 | 1 | 1 | 1 | 1 | 1 | 10 |
| 2 | 1 | 1 | 2 | 2 | 2 | 12 |
| 3 | 1 | 2 | 1 | 2 | 2 | 6 |
| 4 | 1 | 2 | 2 | 1 | 1 | 4 |
| 5 | 2 | 2 | 2 | 1 | 2 | 1 |
| 6 | 2 | 2 | 3 | 2 | 1 | 8 |
| 7 | 2 | 3 | 2 | 2 | 1 | 4 |
| 8 | 2 | 3 | 3 | 1 | 2 | 7 |
| | | | | | Total | 52 |

In Table 14.13, the Idle column corresponds to the first column of the standard $L_8$ OA and columns 2 and 3 corresponds to the columns 2 and 4 of $L_8$ OA respectively. Columns 3 and 5 of the standard $L_8$ OA are dropped. Columns 4 and 5 of Table 14.13 corresponds to the columns 6 and 7 of the standard $L_8$ OA respectively.

***Data analysis from idle column method:*** Consider the response data from one replication given in Table 14.13. The response totals are computed first as usual (Table 14.14).

**TABLE 14.14** Response totals for idle column design

| *Level* | *Factors* | | | *Level* | | *A* | *B* |
|---|---|---|---|---|---|---|---|
| | | | | Idle 1 ($I_1$) | 1 | 22 | 16 |
| | Idle ($I$) | $C$ | $D$ | | 2 | 10 | 16 |
| 1 | 32 | 22 | 26 | Idle 2 ($I_2$) | 2 | 9 | 5 |
| 2 | 20 | 30 | 26 | | 3 | 11 | 15 |

***Computation of sum of squares:***

$$\text{CF} = \frac{(52)^2}{8} = 338.00$$

$$SS_{\text{Total}} = (10)^2 + (12)^2 + \ldots + (7)^2 - \text{CF}$$
$$= 426 - 338 = 88.00$$

The sum of squares of the idle column $SS_I$ is computed from the idle column.

$$SS_I = \frac{(I_1^2 + I_2^2)}{n_i} - \text{CF}$$

where $n_i$ is the number of observations in the total $I_i$

$$= \frac{(32)^2 + (20)^2}{4} - 338$$

$$= 356 - 338 = 18$$

Similarly computing, $SS_C = 8$, and $SS_D = 0$.

The sum of squares for the idle column factors $A$ and $B$ are computed as follows:

$$SS_{A_{1-2}} = \left(\frac{AI_{11}^2}{n_{AI_{11}}}\right) + \left(\frac{AI_{12}^2}{n_{AI_{12}}}\right) - \frac{(AI_{11} + AI_{12})^2}{n_{AI_{11}} + n_{AI_{12}}}$$

$$= \frac{(22)^2}{2} + \frac{(10)^2}{2} - \frac{(22+10)^2}{2+2}$$

$$= 292 - 256 = 36.00$$

$$SS_{A_{2-3}} = \left(\frac{AI_{22}^2}{n_{AI_{22}}}\right) + \left(\frac{AI_{23}^2}{n_{AI_{23}}}\right) - \frac{(AI_{22} + AI_{23})^2}{n_{AI_{22}} + n_{AI_{23}}}$$

$$= \frac{(9)^2}{2} + \frac{(11)^2}{2} - \frac{(9+11)^2}{2+2}$$

$$= 101 - 100 = 1.00$$

Similarly,

$$SS_{B_{1-2}} = \frac{(16)^2}{2} + \frac{(16)^2}{2} - \frac{(16+16)^2}{2+2}$$

$$= 256 - 256 = 0.00$$

$$SS_{B_{2-3}} = \frac{(5)^2}{2} + \frac{(15)^2}{2} - \frac{(15+5)^2}{2+2}$$

$$= 125 - 100 = 25.00$$

The ANOVA summary is given in Table 14.15.

**TABLE 14.15** ANOVA for idle column experiment

| *Source* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ | *Significance* |
|---|---|---|---|---|---|
| Idle | 18.00 | 1 | 18 | 8 | Significant |
| $A_{1-2}$ | 36.00 | 1 | 36 | 16 | Significant |
| $A^*_{2-3}$ | 1.00 | 1 | — | — | |
| $B^*_{1-2}$ | 0.00 | 1 | — | — | |
| $B_{2-3}$ | 25.00 | 1 | 25 | 11.11 | Significant |
| $C^*$ | 8.00 | 1 | — | — | |
| $D^*$ | 0.00 | 1 | — | — | |
| Pooled error | 9.00 | 4 | 2.25 | | |

* Pooled into error, $F_{5\%,1,4} = 7.71$.

From the ANOVA it is seen that the factors $A_{1-2}$ and $B_{2-3}$ are significant. This indicates that the difference in the average condition $A_1$ to $A_2$ and $B_2$ to $B_3$ is significant. The best levels for $A$ and $B$ can be selected by computing the average response for $A_1$, $A_2$, $A_3$ and $B_1$, $B_2$ and $B_3$.

From $I_1, \bar{A}_1 = \frac{22}{2} = 11$ and $\bar{A}_2 = \frac{10}{2} = 5$

From $I_2, \bar{A}_2 = \frac{9}{2} = 4.5$ and $\bar{A}_3 = \frac{11}{2} = 5.5$

Similarly, $\bar{B}_1 = 8$ and $\bar{B}_2 = 8$ (from $I_1$)

$\bar{B}_2 = 2.5$ and $\bar{B}_3 = 7.5$ (from $I_2$)

If the quality characteristic is lower—the better type, the best levels for factors $A$ and $B$ are $A_2$ and $B_2$ respectively. Since factors $C$ and $D$ are not significant, their levels can be selected based on economies and convenience.

**ILLUSTRATION 14.1**

In a study of synthetic rubber manufacturing process, eight factors (*A*, *B*, *C*, *D*, *E*, *F*, *G* and *H*) have been considered. All factors have two-levels except factor *B* which has four-levels. In addition, the experimenter wants to investigate the two-factor interactions *AB*, *AC* and *AG*. Design a matrix experiment.

**Solution:** Required degrees of freedom for the problem are

| | |
|---|---|
| Seven two-level factors: | 7 df |
| One four-level factor (*B*): | 3 df |
| Interaction AB: | 3 df |
| Interaction AC: | 1 df |
| Interaction AG: | 1 df |
| Total degrees of freedom: | 15 |

Hence, $L_{16}$ OA is selected for the design.

The required linear graph for Illustration 14.1 is shown in Figure 14.2.

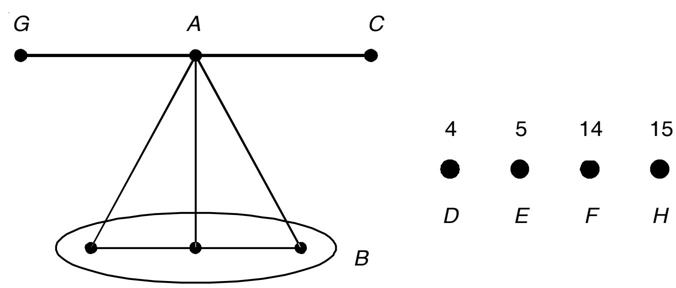



**FIGURE 14.2** Required linear graph for Illustration 14.1.

The required linear graph (Figure 14.2) is superimposed on the standard linear graph and is shown in Figure 14.3. The design layout is given in Table 14.16.

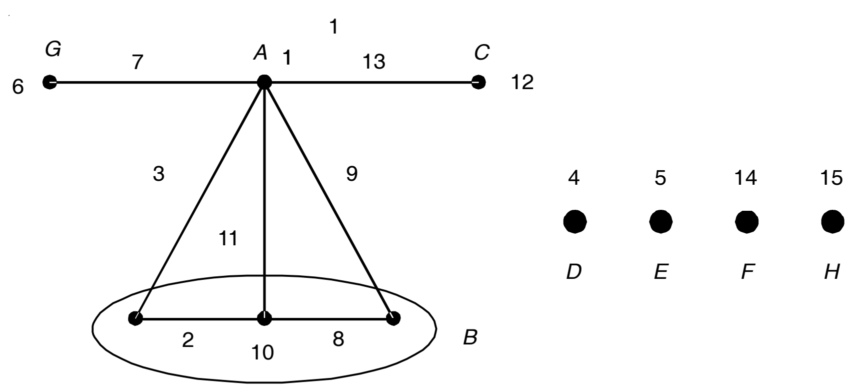



**FIGURE 14.3** Superimposed linear graph for Illustration 14.1.

**TABLE 14.16** Layout of the experimental design for Illustration 14.1

| *Trial* *no.* | *A* 1 | *B* 2,8,10 | *AB* 3 | *D* 4 | *E* 5 | *G* 6 | *AG* 7 | *AB* 9 | *AB* 11 | *C* 12 | *AC* 13 | *F* 14 | *H* 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 |
| 4 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 |
| 5 | 1 | 3 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 |
| 6 | 1 | 4 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 1 |
| 7 | 1 | 3 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 1 |
| 8 | 1 | 4 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 2 |
| 9 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 1 | 2 |
| 10 | 2 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 1 |
| 11 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 1 |
| 12 | 2 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 2 |
| 13 | 2 | 3 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 |
| 14 | 2 | 4 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 |
| 15 | 2 | 3 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 |
| 16 | 2 | 4 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 |

## ILLUSTRATION 14.2

An experimenter wants to study two four-level factors ($P$ and $Q$) and five two-level factors ($A$, $B$, $C$, $D$ and $E$). Further he is also interested to study the two factor interactions $AB$ and $AC$. Design an experiment.

**Solution:** The problem has 13 degrees of freedom. Hence $L_{16}$ OA is required. The standard $L_{16}$ OA has to be modified to accommodate the two four-level factors. The required linear graph is shown in Figure 14.4.

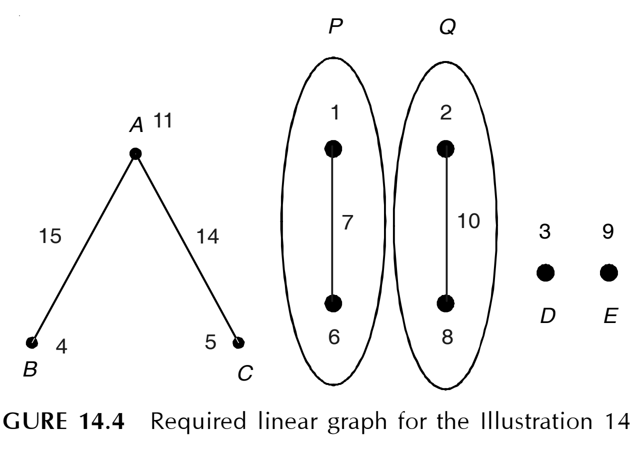



**FIGURE 14.4** Required linear graph for the Illustration 14.2.

The standard linear graph and its modification to suit the required linear graph is shown in Figure 14.5.

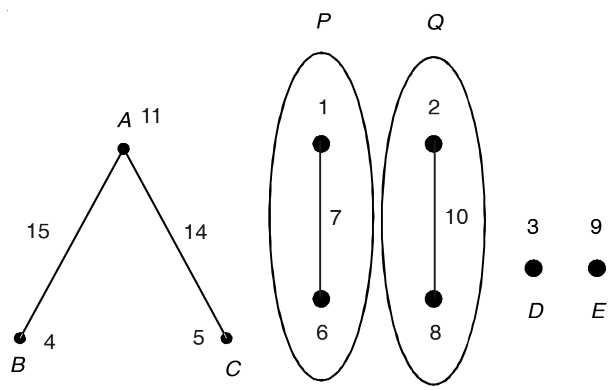



**FIGURE 14.5** Standard linear graph and its modification for Illustration 14.2.

The experimental design with the assignment of factors and interactions is shown in Table 14.17.

**TABLE 14.17** Layout of the experimental design for the Illustration 14.2

| *Trial* *no.* | *P* 1, 6, 7 | *Q* 2, 8, 10 | *D* 3 | *B* 4 | *C* 5 | *E* 9 | *A* 11 | *AC* 14 | *AB* 15 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 |
| 3 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 |
| 4 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 1 |
| 5 | 2 | 3 | 2 | 1 | 1 | 1 | 2 | 2 | 2 |
| 6 | 2 | 4 | 2 | 1 | 1 | 2 | 1 | 1 | 1 |
| 7 | 1 | 3 | 2 | 2 | 2 | 1 | 2 | 1 | 1 |
| 8 | 1 | 4 | 2 | 2 | 2 | 2 | 1 | 2 | 2 |
| 9 | 3 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 2 |
| 10 | 3 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 |
| 11 | 4 | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 1 |
| 12 | 4 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 |
| 13 | 4 | 3 | 1 | 1 | 2 | 2 | 1 | 2 | 1 |
| 14 | 4 | 4 | 1 | 1 | 2 | 1 | 2 | 1 | 2 |
| 15 | 3 | 3 | 1 | 2 | 1 | 2 | 1 | 1 | 2 |
| 16 | 3 | 4 | 1 | 2 | 1 | 1 | 2 | 2 | 1 |

## PROBLEMS

**14.1** An engineer wants to study one three-level factor and four two-level factors. Design an OA experiment.

**14.2** Show how to modify $L_{16}$ OA to accommodate one four-level factor ($A$) and seven two-level factors ($B$, $C$, $D$, $E$, $F$, $G$, and $H$). Suppose the response for the 16 successive trials is

46, 30, 60, 56, 60, 60, 40, 50, 56, 52, 38, 40, 58, 42, 22 and 28.

Analyse the data using ANOVA and draw conclusions

**14.3** Design an OA experiment to study two four-level factors and five two-level factors.

**14.4** An experimenter wants to study three-factors each at three-levels and all two-factor interactions. Design an OA experiment.

**14.5** The following data has been collected (Table 14.18) using the following dummy level design.

**TABLE 14.18** Data for Problem 14.5

| *Trial no.* | *A* | *B* | *C* | *D* | *Response (Y)* | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | $R_1$ | $R_2$ |
| 1 | 1 | 1 | 1 | 1 | 12 | 10 |
| 2 | 1 | 2 | 2 | 2 | 3 | 5 |
| 3 | 1 | 3 | 1′ | 3 | 1 | 2 |
| 4 | 2 | 1 | 2 | 3 | 6 | 8 |
| 5 | 2 | 2 | 1′ | 1 | 11 | 13 |
| 6 | 2 | 3 | 1 | 2 | 6 | 5 |
| 7 | 3 | 1 | 1′ | 2 | 8 | 9 |
| 8 | 3 | 2 | 1 | 3 | 10 | 8 |
| 9 | 3 | 3 | 2 | 1 | 0 | 1 |

Analyse the data using ANOVA and draw conclusions.

# CHAPTER 15 Multi-response Optimization Problems

## 15.1 INTRODUCTION

In Taguchi methods we mainly deal with only single response optimization problems. That is, only one dependent variable (response) is considered and the optimal levels for the parameters are determined based on the mean response/maximum of mean S/N ratio. However, in practice we may have more than one dependant variable. Some of the examples where we have more than one dependent variable or multiple responses are given in Table 15.1.

**TABLE 15.1** Examples of multi-response experiments

| *Experiment* | *Dependant variables/Responses* |
|---|---|
| Heat treatment of machined parts | Surface hardness<br>Depth of hardness |
| Injection moulding of polypropylene components | Tensile strength<br>Surface roughness |
| Optimization of face milling process | Volume of material removal<br>Surface roughness and<br>Height of burr formed |
| Gear hobbing operation | Left profile error<br>Right profile error<br>Left helix error and<br>Right helix error |

Taguchi method cannot be used directly to optimize the multi-response problems. However, we can collect the observed data for each response using Taguchi designs and the data can be analysed by different methods developed by various researchers. In this chapter we will discuss some of the methods which can be easily employed.

Most of the published literature on Taguchi method application deals with a single response. In multi-response problems if we try to determine the optimal levels for the factors based on one response at a time, we may get different set of optimal levels for each response. And it will be difficult to select the best set. Usually, the general approach in these problems is to combine the multi responses into a single statistic (response) and then obtain the optimal levels. Most of the methods available in literature to solve multi-response problems address this issue. Jeyapaul et al. (2005) have presented a literature review on solving multi-response problems in the Taguchi method. In this chapter, the following methods are discussed:

1. Engineering judgment
2. Assignment of weights
3. Data envelopment analysis based ranking (DEAR) approach
4. Grey relational analysis
5. Factor analysis (Principal component method)
6. Genetic algorithm

## 15.2 ENGINEERING JUDGMENT

Engineering judgment has been used until recently to optimize the multi-response problem. Phadke (2008) has used Taguchi method to study the surface defects and thickness of a wafer in the polysilicon deposition process for a VLSI circuit manufacturing. Based on the judgment of relevant experience and engineering knowledge, some trade-offs were made to choose the optimal-factor levels for this two-response problem.

Reddy et al. (1998) in their study in an Indian plastic industry have determined the optimal levels for the control factors considering each response separately. If there is a conflict between the optimum levels arrived at by different response variable, the authors have suggested the use of engineering judgment to resolve the conflict. They have selected a factor that has the smallest or no effect on the S/N ratio for all the response variables but has a significant effect on the mean levels and termed it *mean adjustment/signal factor*. They have set the level of the adjustment factor so that the mean response is on target.

By human judgment, validity of experimental results cannot be easily assured. Each experimenter can judge differently. Also engineering judgment together with past experience will bring in some uncertainty in the decision making process.

## 15.3 ASSIGNMENT OF WEIGHTS

In assignment of weights method, the multi-response problem is converted into a single response problem. Suppose we have two responses in a problem. Let $W_1$ be the weight assigned to, say the first response $R_1$ and $W_2$ be the weight assigned to the second response $R_2$. The sum of the weighted response ($W$) will be the single response, where

$$W = W_1R_1 + W_2R_2 \tag{15.1}$$

This ($W$) is termed Multi Response Performance Index (MRPI). Using this MRPI, the problem is solved as a single response problem. In the multi-response problem, each response can be the original observed data or its transformation such as S/N ratio. In this approach, the major issue is the method of determining the weights. Literature review indicates that several approaches have been used to obtain MRPI.

## ILLUSTRATION 15.1

In an injection moulding process, the components were rejected due to defects such as tearing along the hinge and poor surface finish. It was identified that tensile strength and surface roughness have been the causes for the defects. The following factors and levels were selected for study [Table 15.2(a)].

**TABLE 15.2(a)** Factors and levels for Illustration 15.1

| *Factors* | *Levels* | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| Inlet temperature ($A$) | 250 | 265 | 280 |
| Injection time ($B$) | 3 | 6 | 9 |
| Injection pressure ($C$) | 30 | 55 | 80 |

The experiment results are given in Table 15.2(b). Note that there are two responses. One is the Tensile Strength (TS) and the second is the Surface Roughness (SR). Also note that TS is larger—the better type of quality characteristic and SR is smaller—the better type of quality characteristic.

**TABLE 15.2(b)** Experimental results for Illustration 15.1

| *Trial no.* | *Factors* A | *B* | *C* | *TS* | *SR* |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1075 | 0.3892 |
| 2 | 1 | 2 | 2 | 1044 | 0.3397 |
| 3 | 1 | 3 | 3 | 1062 | 0.6127 |
| 4 | 2 | 1 | 2 | 1036 | 0.9640 |
| 5 | 2 | 2 | 3 | 988 | 0.4511 |
| 6 | 2 | 3 | 1 | 985 | 0.3736 |
| 7 | 3 | 1 | 3 | 926 | 1.2710 |
| 8 | 3 | 2 | 1 | 968 | 1.2910 |
| 9 | 3 | 3 | 2 | 957 | 0.1577 |

The weights are determined as follows. For TS (larger—the better characteristic), the individual response (data) is divided by the total response value ($\Sigma$ TS). In the case of SR (smaller—the better characteristic), reverse normalization procedure is used. That is, for each response data, 1/SR is obtained and then $W_{SR}$ is computed. From Table 15.2(b), $\Sigma$ TS = 9041 and $\Sigma$ 1/SR = 20.9785.

Note that
$$W_{TS_1} = \frac{TS_1}{\Sigma TS} \text{ and } W_{SR_1} = \frac{(1/SR_1)}{\Sigma 1/SR}$$

For example, in the first trial $W_{TS_1} = \dfrac{1075}{9031} = 0.1190$ and $W_{SR_1} = \dfrac{2.5694}{20.9785} = 0.1225$

$$(\text{MRPI})_i = W_1Y_{11} + W_2Y_{12} + \cdots + W_jY_{ij} \tag{15.2}$$

$(\text{MRPI})_i$ = MRPI of the $i$th trial/experiment
$W_j$ = Weight of the $j$th response/dependant variable
$Y_{ij}$ = Observed data of $i$th trial/experiment under $j$th response

$$\text{MRPI}_1 = (0.1189 \times 1075 + 0.1225 \times 0.3892) = 127.8652$$

The weights and MRPI values for all the trials are given in Table 15.3.

**TABLE 15.3** Weights and MRPI values for Illustration 15.1

| *Trial* | TS | $W_{TS}$ | SR | 1/SR | $W_{SR}$ | MRPI |
|---|---|---|---|---|---|---|
| 1 | 1075 | 0.1189 | 0.3892 | 2.5694 | 0.1225 | 127.8652 |
| 2 | 1044 | 0.1155 | 0.3397 | 2.9438 | 0.1403 | 120.6297 |
| 3 | 1062 | 0.1175 | 0.6127 | 1.6321 | 0.0778 | 124.8367 |
| 4 | 1036 | 0.1145 | 0.9640 | 1.0373 | 0.0494 | 118.6696 |
| 5 | 988 | 0.1093 | 0.4511 | 2.2168 | 0.1057 | 108.0356 |
| 6 | 985 | 0.1090 | 0.3736 | 2.6767 | 0.1276 | 107.4127 |
| 7 | 926 | 0.1024 | 1.2710 | 0.7867 | 0.0375 | 94.8701 |
| 8 | 968 | 0.1071 | 1.2910 | 0.7746 | 0.0369 | 103.7204 |
| 9 | 957 | 0.1058 | 0.1577 | 6.3411 | 0.3023 | 101.2983 |

Now, we consider MRPI (Table 15.4) as a single response of the original problem and obtain solution using methods discussed in Chapter 12. Since MRPI is a weighted score, optimal levels are identified based on maximum MRPI values. The original problem with the MRPI score is given in Table 15.4.

**TABLE 15.4** MRPI as response (Illustration 15.1)

| *Trial* | *Factors* | | | |
|---|---|---|---|---|
| *no.* | *A* | *B* | *C* | MRPI |
| 1 | 1 | 1 | 1 | 127.8652 |
| 2 | 1 | 2 | 2 | 120.6297 |
| 3 | 1 | 3 | 3 | 124.8367 |
| 4 | 2 | 1 | 2 | 118.6696 |
| 5 | 2 | 2 | 3 | 108.0356 |
| 6 | 2 | 3 | 1 | 107.4127 |
| 7 | 3 | 1 | 3 | 94.8701 |
| 8 | 3 | 2 | 1 | 103.7204 |
| 9 | 3 | 3 | 2 | 101.2983 |

The level totals of MRPI are given in Table 15.5.

**TABLE 15.5** Level totals of MRPI for Illustration 15.1

| *Factors* | *Levels* | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| Inlet temperature ($A$) | **373.3316** | 334.1179 | 299.8888 |
| Injection time ($B$) | **341.4049** | 332.3857 | 333.5477 |
| Injection pressure ($C$) | 338.9983 | **340.5976** | 327.7424 |

The optimal levels are selected based on maximum MRPI are $A_1$, $B_1$ and $C_2$.

## ILLUSTRATION 15.2

Suppose we have the following Taguchi design with data (Table 15.6) on the following three responses/dependent variables from a manufacturing operation:

1. Volume of material removal (Larger—the better type quality characteristic)
2. Surface roughness (Smaller—the better type quality characteristic)
3. Burr height (Smaller—the better type quality characteristic)

Note that there are two replications under each response.

**TABLE 15.6** Experimental results for Illustration 15.2

| *Trial no.* | *Factors* | | | *Material Removal (MR)* | | *Surface Roughness (SR)* | | *Burr Height (BH)* | |
|---|---|---|---|---|---|---|---|---|---|
| | *A* | *B* | *C* | 1 | 2 | 1 | 2 | 1 | 2 |
| 1 | 1 | 1 | 1 | 160 | 110 | 7.8 | 8.3 | 4.15 | 4.50 |
| 2 | 1 | 2 | 2 | 151 | 145 | 7.0 | 7.5 | 3.20 | 3.27 |
| 3 | 1 | 3 | 3 | 152 | 120 | 9.2 | 9.6 | 3.21 | 4.73 |
| 4 | 2 | 1 | 2 | 111 | 149 | 5.8 | 4.6 | 3.60 | 3.79 |
| 5 | 2 | 2 | 3 | 95 | 140 | 4.7 | 5.8 | 2.50 | 3.00 |
| 6 | 2 | 3 | 1 | 112 | 91 | 8.2 | 6.2 | 3.78 | 3.67 |
| 7 | 3 | 1 | 3 | 97 | 77 | 4.8 | 6.2 | 2.93 | 2.94 |
| 8 | 3 | 2 | 1 | 77 | 79 | 4.6 | 3.8 | 2.84 | 2.53 |
| 9 | 3 | 3 | 2 | 97 | 81 | 5.5 | 5.7 | 2.98 | 2.98 |

The data analysis using weightage method is presented below. Since two replications are available, the data is transformed into the S/N ratios and is given in Table 15.7.

The weights are determined as explained in Illustration 15.1. For MR, the individual response is divided by the total response value ($\Sigma$ MR). In the case of SR and BH, reverse normalization procedure is used. While computing the weights the absolute value of data is considered. The weights obtained and the MRPI values are given in Table 15.8. MRPI is computed using Eq. (15.2).

**TABLE 15.7** S/N ratio values for Illustration 15.2

| *Trial no.* | *S/N ratio values* | | |
|---|---|---|---|
| | *MR* | *SR* | *BH* |
| 1 | 42.1574 | –18.1201 | –12.7268 |
| 2 | 43.3998 | –17.2119 | –10.1980 |
| 3 | 42.4900 | –19.4645 | –12.1321 |
| 4 | 41.9995 | –14.3775 | –11.3552 |
| 5 | 40.9198 | –14.4506 | –8.8224 |
| 6 | 39.9896 | –17.2296 | –11.4235 |
| 7 | 38.6178 | –14.8770 | –9.3522 |
| 8 | 37.8397 | –12.5042 | –8.5933 |
| 9 | 38.8824 | –14.9651 | –9.4843 |

**TABLE 15.8** Weights and MRPI for Illustration 15.2

| *Trial no.* | *Weights* | | | *MRPI* $\times 10^3$ |
|---|---|---|---|---|
| | $W_{MV}$ | $W_{SR}$ | $W_{BH}$ | |
| 1 | 0.1151 | 0.0960 | 0.0896 | 2.0060 |
| 2 | 0.1185 | 0.1010 | 0.1118 | 2.2643 |
| 3 | 0.1159 | 0.0894 | 0.0940 | 2.0440 |
| 4 | 0.1147 | 0.1209 | 0.1005 | 1.9379 |
| 5 | 0.1117 | 0.1203 | 0.1292 | 1.6858 |
| 6 | 0.1092 | 0.1009 | 0.0998 | 1.4883 |
| 7 | 0.1054 | 0.1169 | 0.1297 | 1.1905 |
| 8 | 0.1033 | 0.1391 | 0.1328 | 1.0283 |
| 9 | 0.1061 | 0.1162 | 0.1202 | 1.2465 |

The optimal levels for the factors can now be determined considering MRPI as single response as in Illustration 15.1. This is left as an exercise to the reader.

## 15.4 DATA ENVELOPMENT ANALYSIS BASED RANKING METHOD

Hung-Chang and Yan-Kwang (2002) proposed a Data Envelopment Analysis based Ranking (DEAR) method for optimizing multi-response Taguchi experiments. In this method, a set of original responses are mapped into a ratio (weighted sum of responses with larger—the better is divided by weighted sum of responses with smaller—the better or nominal—the best) so that the optimal levels can be found based on this ratio. This ratio can be treated as equivalent to MRPI.

The following steps discuss DEAR:

*Step* 1: Determine the weights associated with each response for all experiments using an appropriate weighting technique.

*Step* 2: Transform the observed data of each response into weighted data by multiplying the observed data with its own weight.

*Step* 3: Divide the weighted data of larger—the better type with weighted data of smaller—the better type or nominal—the best type.

*Step* 4: Treat the value obtained in Step 3 as MRPI and obtain the solution.

## ILLUSTRATION 15.3

The problem considered here is same as in Illustration 15.1. For determining the weights, the application of the first step in this approach results in Table 15.8(a). Now applying Step 2 and Step 3, we obtain Table 15.8(b). Note that the MRPI values are obtained by dividing the weighted response of larger—the better (TS) with the weighted response of smaller—the better quality characteristic.

**TABLE 15.8(a)** Response weights for Illustration 15.3

| *Trial no.* | *TS* | $W_{TS}$ | *SR* | $W_{SR}$ |
|---|---|---|---|---|
| 1 | 1075 | 0.1189 | 0.3892 | 0.1225 |
| 2 | 1044 | 0.1155 | 0.3397 | 0.1403 |
| 3 | 1062 | 0.1175 | 0.6127 | 0.0778 |
| 4 | 1036 | 0.1145 | 0.9640 | 0.0494 |
| 5 | 988 | 0.1093 | 0.4511 | 0.1057 |
| 6 | 985 | 0.1090 | 0.3736 | 0.1276 |
| 7 | 926 | 0.1024 | 1.2710 | 0.0375 |
| 8 | 968 | 0.1071 | 1.2910 | 0.0369 |
| 9 | 957 | 0.1058 | 0.1577 | 0.3023 |

**TABLE 15.8(b)** Weighted responses and MRPI for Illustration 15.3

| *Trial no.* | TS * $W_{TS}$ ($P$) | SR * $W_{SR}$($Q$) | MRPI = $P/Q \times 10^3$ |
|---|---|---|---|
| 1 | 127.8175 | 0.04768 | 2.6807 |
| 2 | 120.5820 | 0.04766 | 2.5300 |
| 3 | 124.7850 | 0.04767 | 2.6177 |
| 4 | 118.6220 | 0.04762 | 2.4910 |
| 5 | 107.9884 | 0.04768 | 2.2648 |
| 6 | 107.3650 | 0.04767 | 2.2518 |
| 7 | 94.8224 | 0.04766 | 1.9896 |
| 8 | 103.6728 | 0.04764 | 2.1762 |
| 9 | 101.2506 | 0.04767 | 2.1240 |

The optimal levels are identified treating MRPI as single response as in Illustration 15.1. The level totals of MRPI for Illustration 15.3 are given in Table 15.9. The optimal levels based on maximum MRPI are $A_1$, $B_1$ and $C_2$.

**TABLE 15.9** Level totals of MRPI for Illustration 15.3 (x$10^3$)

| *Factors* | *Levels* | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| Inlet temperature (*A*) | **7.8284** | 7.0076 | 6.2898 |
| Injection time (*B*) | **7.1613** | 6.9710 | 6.9935 |
| Injection pressure (*C*) | 7.1087 | **7.1450** | 6.8721 |

## ILLUSTRATION 15.4

Application of DEAR approach for Illustration 15.2.

The weights obtained for Illustration 15.2 are reproduced in Table 15.10. Now the weighted response for all the experiments is obtained as in Illustration 15.3 and is given in Table 15.10. The MRPI values given in Table 5.10 are obtained by dividing the weighted response of the larger—the better type with the sum of weighted response of the smaller—the better type of quality characteristic. Now considering MRPI as single response, the MRPI data can be analysed and optimal levels for the factors can be determined.

**TABLE 15.10** Weights, weighted response and MRPI for Illustration 15.4

| *Trial no.* | *Weights* | | | MV * $W_{MV}$ | SR * $W_{SR}$ | BH * $W_{BH}$ | MRPI = |
|---|---|---|---|---|---|---|---|
| | $W_{MV}$ | $W_{SR}$ | $W_{BH}$ | (*P*) | (*Q*) | (*R*) | *P*/(*Q* + *R*) |
| 1 | 0.1151 | 0.0960 | 0.0896 | 31.0770 | 1.5456 | 0.7750 | 13.3918 |
| 2 | 0.1185 | 0.1010 | 0.1118 | 35.0760 | 1.4645 | 0.7233 | 16.0325 |
| 3 | 0.1159 | 0.0894 | 0.0940 | 31.5480 | 1.6807 | 0.7746 | 12.8395 |
| 4 | 0.1147 | 0.1209 | 0.1005 | 29.8220 | 1.2574 | 0.7427 | 14.9102 |
| 5 | 0.1117 | 0.1203 | 0.1292 | 26.2495 | 1.2631 | 0.7106 | 13.2996 |
| 6 | 0.1092 | 0.1009 | 0.0998 | 22.1676 | 1.4539 | 0.7435 | 10.0881 |
| 7 | 0.1054 | 0.1169 | 0.1297 | 18.3396 | 1.2859 | 0.7155 | 9.1634 |
| 8 | 0.1033 | 0.1391 | 0.1328 | 16.1148 | 1.1684 | 0.7131 | 8.5649 |
| 9 | 0.1061 | 0.1162 | 0.1202 | 18.8858 | 1.3014 | 0.7164 | 9.3596 |

## 15.5 GREY RELATIONAL ANALYSIS

Grey relational analysis is used for solving interrelationships among the multiple responses. In this approach a grey relational grade is obtained for analysing the relational degree of the multiple responses. Lin et al. (2002) have attempted grey relational based approach to solve multi-response problems in the Taguchi methods.

### *Optimization steps in grey relational analysis*

*Step* 1: Transform the original response data into S/N ratio ($Y_{ij}$) using the appropriate formulae depending on the type of quality characteristic.

*Step* 2: Normalize $Y_{ij}$ as $Z_{ij}$ ($0 \le Z_{ij} \le 1$) by the following formula to avoid the effect of using different units and to reduce variability. Normalization is a transformation performed on a single input to distribute the data evenly and scale it into acceptable range for further analysis.

$Z_{ij}$ = Normalized value for $i$th experiment/trial for $j$th dependant variable/response

$$Z_{ij} = \frac{Y_{ij} - \min(Y_{ij}, i = 1, 2, ..., n)}{\max(Y_{ij}, i = 1, 2, ..., n) - \min(Y_{ij}, i = 1, 2, ..., n)} \tag{15.3}$$

(to be used for S/N ratio with larger—the better case)

$$Z_{ij} = \frac{\max(Y_{ij}, i = 1, 2, ..., n) - Y_{ij}}{\max(Y_{ij}, i = 1, 2, ..., n) - \min(Y_{ij}, i = 1, 2, ..., n)} \tag{15.4}$$

(to be used for S/N ratio with smaller—the better case)

$$Z_{ij} = \frac{(|Y_{ij} - T|) - \min(|Y_{ij} - T|, i = 1, 2, ..., n)}{\max(|Y_{ij} - T|, i = 1, 2, ..., n) - \min(|Y_{ij} - T|, i = 1, 2, ..., n)} \tag{15.5}$$

(to be used for S/N ratio with nominal—the best case)

*Step* 3: Compute the grey relational coefficient (GC) for the normalized S/N ratio values.

$$\mathrm{GC}_{ij} = \frac{\Delta_{\min} + \lambda\Delta_{\max}}{\Delta_{ij} + \lambda\Delta_{\max}} \quad \begin{cases} i = 1, 2, ..., n\text{—experiments} \\ j = 1, 2, ..., m\text{—responses} \end{cases} \tag{15.6}$$

$\mathrm{GC}_{ij}$ = grey relational coefficient for the $i$th experiment/trial and $j$th dependant variable/response

$\Delta$ = absolute difference between $Y_{oj}$ and $Y_{ij}$ which is a deviation from target value and can be treated as quality loss.

$Y_{oj}$ = optimum performance value or the ideal normalized value of $j$th response

$Y_{ij}$ = the $i$th normalized value of the $j$th response/dependant variable

$\Delta_{\min}$ = minimum value of $\Delta$

$\Delta_{\max}$ = maximum value of $\Delta$

$\lambda$ is the distinguishing coefficient which is defined in the range $0 \le \lambda \le 1$ (the value may be adjusted on the practical needs of the system)

*Step* 4: Compute the grey relational grade ($G_i$).

$$G_i = \frac{1}{m} \sum \mathrm{GC}_{ij} \tag{15.7}$$

where $m$ is the number of responses

*Step* 5: Use response graph method or ANOVA and select optimal levels for the factors based on maximum average $G_i$ value.

## ILLUSTRATION 15.5

**Application of Grey Relational Analysis**

The problem considered here is same as in Illustration 15.1. However, the data is different in this example which is given in Table 15.11. Note that we had data from one replication in Illustration 15.1 as against two replications in this illustration.

**TABLE 15.11** Experimental data for Illustration 15.5

| *Trial* | *Factors* | | | *Tensile strength* (TS) | | *Surface roughness* (SR) | |
|---|---|---|---|---|---|---|---|
| *no.* | *A* | *B* | *C* | 1 | 2 | 1 | 2 |
| 1 | 1 | 1 | 1 | 1075 | 1077 | 0.3892 | 0.3896 |
| 2 | 1 | 2 | 2 | 1044 | 1042 | 0.3397 | 0.3399 |
| 3 | 1 | 3 | 3 | 1062 | 1062 | 0.6127 | 0.6127 |
| 4 | 2 | 1 | 2 | 1036 | 1032 | 0.9640 | 0.9620 |
| 5 | 2 | 2 | 3 | 988 | 990 | 0.4511 | 0.4515 |
| 6 | 2 | 3 | 1 | 985 | 983 | 0.3736 | 0.3736 |
| 7 | 3 | 1 | 3 | 926 | 926 | 1.2712 | 1.2710 |
| 8 | 3 | 2 | 1 | 968 | 966 | 1.2910 | 1.2908 |
| 9 | 3 | 3 | 2 | 957 | 959 | 0.1577 | 0.1557 |

*Step* 1: Transformation of data into S/N ratios

There are two replications for each response. Now the data has to be transformed into S/N ratio. Since tensile strength (TS) has to be maximized, it is larger—the better type of quality characteristic. Hence the S/N ratio for TS is computed from the following formula:

$$\text{S/N ratio}(\eta) = -10\ \log\left(\frac{1}{n}\right)\sum_{i=1}^{n}\frac{1}{Y_{ij}^2} \tag{15.8}$$

The second performance measure (response) is smaller—the better type and its S/N ratio is computed from

$$\text{S/N ratio}(\eta) = -10\ \log\left(\frac{1}{n}\sum_{i-1}^{n}Y_{ij}^2\right) \tag{15.9}$$

for Trial no. 1,

$$\text{S/N ratio for TS} = -10\ \log\ \frac{1}{2}\left[\frac{1}{(1075)^2}+\frac{1}{(1077)^2}\right] = 60.63624543$$

$$\text{S/N ratio for SR} = -10\ \log\ \frac{1}{2}\ [(0.3892)^2 + (0.3896)^2] = 8.1920799$$

Similarly, for all trials the S/N ratios are computed and given in Table 15.12.

*Step* 2: Normalization of S/N values

Using Eqs. (15.3) and (15.4) we normalize the S/N values in Table 15.12.

For TS, Trial no. 1, min $Y_{ij}$ = 59.3322 (Trial no. 7) and max $Y_{ij}$ = 60.6362 (Trial no. 1).

Applying Eq. 15.3, for Trial no. 1,

$$Z_{11} = \frac{(60.6362\ -\ 59.3322)}{(60.6362\ -\ 59.3322)} = 1$$

**TABLE 15.12** S/N ratios for Illustration 15.5

| *Trial no.* | *Tensile strength* (TS) 1 | 2 | *Surface roughness* (SR) 1 | 2 | *S/N ratio* (TS) | *S/N ratio* (SR) |
|---|---|---|---|---|---|---|
| 1 | 1075 | 1077 | 0.3892 | 0.3896 | 60.6362 | 8.1921 |
| 2 | 1044 | 1042 | 0.3397 | 0.3399 | 60.3657 | 9.3755 |
| 3 | 1062 | 1062 | 0.6127 | 0.6127 | 60.5225 | 4.2550 |
| 4 | 1036 | 1032 | 0.9640 | 0.9620 | 60.2904 | 0.3275 |
| 5 | 988 | 990 | 0.4511 | 0.4515 | 59.9039 | 6.9107 |
| 6 | 985 | 983 | 0.3736 | 0.3736 | 59.8599 | 8.5519 |
| 7 | 926 | 926 | 1.2712 | 1.2710 | 59.3322 | –2.0836 |
| 8 | 968 | 966 | 1.2910 | 1.2908 | 59.7085 | –2.2178 |
| 9 | 957 | 959 | 0.1577 | 0.1557 | 59.6273 | 16.1542 |

and $$Z_{21} = \frac{(60.3657 - 59.3322)}{(60.6362 - 59.3322)} = 0.7925$$

Similarly, the normalized values for TS are obtained and tabulated in Table 15.13.
For SR, min $Y_{ij}$ = –2.2178 (Trial no. 8), max $Y_{ij}$ = 16.1542 (Trial no. 9).

Applying Eq. (15.4); $$Z_{82} = \frac{16.1542 - (-2.2178)}{16.1542 - (-2.2178)} = 1$$

and $$Z_{12} = \frac{16.1542 - (8.1921)}{16.1542 - (-2.2178)} = 0.4334$$

Similarly, the normalized score for other trials of SR are computed and tabulated in Table 15.12.

**TABLE 15.13** Normalized S/N ratios for Illustration 15.5

| *Trial no.* | *S/N ratios* TS | SR | *Normalized S/N ratios* TS | SR |
|---|---|---|---|---|
| 1 | 60.6362 | 8.1921 | 1 | 0.4334 |
| 2 | 60.3657 | 9.3755 | 0.7925 | 0.3369 |
| 3 | 60.5225 | 4.2550 | 0.9128 | 0.6477 |
| 4 | 60.2904 | 0.3275 | 0.7348 | 0.8615 |
| 5 | 59.9039 | 6.9107 | 0.4384 | 0.5031 |
| 6 | 59.8599 | 8.5519 | 0.4047 | 0.4138 |
| 7 | 59.3322 | –2.0836 | 0 | 0.9927 |
| 8 | 59.7085 | –2.2178 | 0.2886 | 1 |
| 9 | 59.6273 | 16.1542 | 0.2263 | 0 |

*Step* 3: Computation of grey relational coefficient ($GC_{ij}$)
Quality loss ($\Delta$) = absolute difference between $Y_{oj}$ and $Y_{ij}$
From Table 15.13, $Y_{oj}$ = 1 for both TS and SR.
For Trial no. 1 and TS; $\Delta_{TS_1} = 1 - 1 = 0$ and for Trial no. 2, $\Delta_{TS_2} = 1 - 0.7925 = 0.2075$.

Similarly, we can compute the Δ values for all the trials and responses. These are given in Table 15.14.

The grey relational coefficient ($GC_{ij}$) is computed from the Eq. (15.6). Assume $\lambda = 1$.

$$GC_{ij} = \frac{\Delta_{min} + \lambda\Delta_{max}}{\Delta_{ij} + \lambda\Delta_{max}}$$

For Illustration 15.5, from Table 15.14, for both responses (TS and SR), $\Delta_{max} = 1$ and $\Delta_{min} = 0$.

For Trial no. 1 and TS, the $GC_{TS_1} = \dfrac{0 + 1 \times 1}{0 + 1 \times 1} = 1$.

For Trial no. 2 of TS, $GC_{TS_2} = \dfrac{0 + 1 \times 1}{0.2075 + 1 \times 1} = 0.8282$. Similarly, the coefficients are computed for all the trials of both TS and SR and are tabulated in Table 15.14.

The grey grade values are computed from Eq. (15.7). For Illustration 15.5, for Trial no. 1,

$$G_i = \frac{1}{m}\sum GC_{ij} = \frac{1}{2}(1 + 0.6383) = 0.8192$$

Similarly for all the trials, the grey grade values are computed and are given in Table 15.14.

**TABLE 15.14** Normalized S/N ratios for Illustration 15.5

| *Trial no.* | *Normalized S/N ratios* TS | TR | $\Delta_{TS}$ | $\Delta_{SR}$ | $GC_{TS}$ | $GC_{SR}$ | $G_i$ |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 0.4334 | 0 | 0.5666 | 1 | 0.6383 | 0.8192 |
| 2 | 0.7925 | 0.3369 | 0.2075 | 0.6631 | 0.8282 | 0.6131 | 0.7206 |
| 3 | 0.9128 | 0.6477 | 0.0872 | 0.3523 | 0.9198 | 0.7395 | 0.8296 |
| 4 | 0.7348 | 0.8615 | 0.2652 | 0.1385 | 0.7904 | 0.8783 | 0.8344 |
| 5 | 0.4384 | 0.5031 | 0.5616 | 0.4969 | 0.6404 | 0.6681 | 0.6542 |
| 6 | 0.4047 | 0.4138 | 0.5953 | 0.5862 | 0.6268 | 0.6304 | 0.6286 |
| 7 | 0 | 0.9927 | 1 | 0.0073 | 0.5 | 0.9927 | 0.7464 |
| 8 | 0.2886 | 1 | 0.7114 | 0 | 0.5843 | 1 | 0.7922 |
| 9 | 0.2263 | 0 | 0.7737 | 1 | 0.5638 | 0.5 | 0.5319 |

Now, the grey grade ($G_i$) is equivalent to MRPI and is treated as single response problem and MRPI data is analysed to determine the optimal levels for the factors.

The main effects on MRPI (mean of MRPI) are tabulated in Table 15.15.

**TABLE 15.15** Mean of MRPI values for Illustration 15.5

| *Factors* | *Levels* 1 | 2 | 3 |
|---|---|---|---|
| Inlet temperature (*A*) | **0.7898** | 0.7057 | 0.6902 |
| Injection time (*B*) | **0.8000** | 0.7223 | 0.6634 |
| Injection pressure (*C*) | **0.7466** | 0.6956 | 0.7434 |

From Table 15.15, it is found that the optimal levels are $A_1$, $B_1$ and $C_1$.

## 15.6 FACTOR ANALYSIS

Factor analysis is the most often used multi-variate technique in research studies. This technique is applicable when there is a systematic interdependence among a set of observed variables. Factor analysis aims at grouping the original input variables into factors which underlie the input variables. Each factor is accounted for one or more input variables. Principal Component Method (PCM) is a mathematical procedure that summarizes the information contained in a set of original variables into a new and smaller set of uncorrelated combinations with a minimum loss of information. This analysis combines the variables that account for the largest amount of variance to form the first principal component. The second principle component accounts for the next largest amount of variance, and so on, until the total sample variance is combined into component groups.

PCM of factor analysis is used for solving the multi-response Taguchi problems. Chao-Ton and Lee-Ing Tong (1997) have developed an effective procedure to transform a set of responses into a set of uncorrelated components such that the optimal conditions in the parameter design stage for the multi-response problem can be determined. They have proposed an effective procedure on the basis of PCM to optimize the multi-response problems in the Taguchi method.

Usually, Statistical software is like SPPSS, MINITAB, etc. are used to perform factor analysis. For details on factor analysis, the reader may refer to Research Methodology by Panneerselvam (2004).

The procedure for application of factor analysis is discussed as follows:

*Step* 1: Transform the original data from the Taguchi experiment into S/N ratios for response using the appropriate formula depending on the type of quality characteristic (Section 13.3).

*Step* 2: Normalize the S/N ratios as in the case of Illustration 15.5.

*Step* 3: The normalized S/N ratio values corresponding to each response are considered to be the initial input for Factor Analysis.

*Step* 3.1: Input the data matrix of $m \times n$ size, where $m$ is the number of sets of observations and $n$ is the number of characteristics (input variables). Input the total number of principal components ($K$) to be identified.

*Step* 3.2: Find $n \times n$ correlation coefficient matrix $R_1$ summarizing the correlation coefficient for each pair of variables. Each diagonal value of the correlation coefficient matrix is assumed as 1, since the respective variable is correlated to itself.

*Step* 3.3: Perform reflections if necessary:

3.3.1: If the matrix is positive manifold, then assign a weight ($w_j$) of +1 to each column $j$ (variable $j$) in the matrix; then, treat the matrix $R_k$ as the matrix $R_k$ without any modification in $R_k$ and go to Step 3.4; otherwise go to Step 3.3.2.

3.3.2: Perform reflections:

(a) Initially assign a weight ($w_j$) of +1 for all the columns.

(b) If there is a negative entry in column $j$, it should be reflected; assign –1 as the weight for that column $j$.

In this case, for each negative value in the row $i$ of the column $j$, do the following updation:

$R_k(i, j)$ = Absolute value of $R_k(i, j)$
$R_k(j, i)$ = Absolute value of $R_k(j, i)$

(c) Treat this revised matrix as the matrix $R'_k$ which is known as the matrix with reflections.

*Step* 3.4: Determination of $k$th principal component:

3.4.1: For each column $j$ in the matrix $R'_k$, find the sum of the entries in it ($S_j$).

3.4.2: Find the grand total ($T_1$), which is the sum of squares of the column sums:

$$T_1 = \sum_{j=1}^{n} S_j^2$$

3.4.3: Find the normalizing factor, $W$ which is given by the formula:

$$W = \sqrt{T_1}$$

3.4.4: Divide each column sum, $S_j$ by $W$ to get unweighted normalized loading $UL_j$ for that column $j$.

3.4.5: For each column $j$, find its weighted factor loading $L_j$ by multiplying $UL_j$ with its weight $w_j$.

3.4.6: Set row number $i$ of the matrix $R_k$ to 1($i$ = 1)

3.4.7: Multiply the value of $L_j$ with the corresponding value in the $i$th row of $R_k$ matrix for each $j$ and sum such products and treat it as $M_i$, where

$$M_i = \sum_{j=1}^{n} R_k(i, j) L_j$$

3.4.8: Set $i = i+1$.

3.4.9: If $i \leq n$, go to Step 3.4.7; otherwise go to Step 3.4.10.

3.4.10: Find the grand total ($T_2$), which is the sum of squares of $M_j$, where $j$ = 1, 2, 3, …, $n$ as

$$T_2 = \sum_{j=1}^{n} M_j^2$$

3.4.11: Find the normalizing factor $N$ which is given by the formula:

$$N = \sqrt{T_2}$$

3.4.12: Divide each $M_j$ by $N$ to get normalized loadings $Q_j$, where

$$Q_j = \frac{M_j}{N}$$

3.4.13: For each $j$, compare $L_j$ and $Q_j$ towards convergence. If all $L_j$ and $Q_j$ are more or less closer, then go to Step 3.5; otherwise go to Step 3.4.14.

3.4.14: Update $L_j = Q_j$, for all $j$ = 1, 2, 3, …, $n$, otherwise go to Step 3.4.6.

*Step* 3.5: Store the loadings of the $k$th principal components $P_{ik}$, as shown below. Latest normalizing factor for $Q$ in Step 3.4.11 of the last iteration of convergence = $N$. Then,

$$P_{ik} = L_i \sqrt{N}$$

*Step* 3.6: Increment $k$ by 1 ($k = k + 1$)

*Step* 3.7: If $k > K$ then go to Step 3.10.

*Step* 3.8: Determination of $R_k$:

3.8.1: Determination of cross product matrix ($C_k$): Obtain the product of each pair of factor loadings of ($k - 1$)th principal component and store it in the respective row and column of the cross product matrix ($C_k$).

3.8.2: Find the residual matrix $R_k$ by element-by-element subtraction of the cross product matrix $C_k$ from the matrix $R_{k-1}$.

*Step* 3.9: Go to Step 3.3.

*Step* 3.10: Arrange the loadings of $K$ principal components by keeping the principal components on columns and the variables in rows.

*Step* 3.11: Find the sum of squares of loadings of each column $j$ (principal component $j$) which is known as eigenvalue of that column $j$.

*Step* 3.12: Drop insignificant principal components which have eigenvalues less than one.

*Step* 3.13: The principal component values with eigenvalue greater than one corresponding to each response is considered for further analysis. The option of performing factor analysis is available in statistical software (SPSS, MINITAB, etc.).

*Step* 4: Calculate Multi Response Performance Index (MRPI) value using principal components obtained by FA.

$$\text{MRPI}_i = P_1Y_{11} + P_2Y_{12} + \cdots + P_jY_{ij} \tag{15.10}$$

*Step* 5: Determine the optimal factor and its level combination.

The higher performance index implies the better product quality, therefore, on the basis of performance index, the factor effect can be estimated and the optimal level for each controllable factor can also be determined. For Example, to estimate the effect of factor $i$, calculate the average of multi-response performance index values (MRPI) for each level $j$, denoted as $\text{MRPI}_{ij}$, then the effect, $E_i$, is defined as:

$$E_i = \max(\text{MRPI}_{ij}) - \min(\text{MRPI}_{ij}) \tag{15.11}$$

If the factor $i$ is controllable, the best level $j^*$, is determined by

$$j^* = \max_j(\text{MRPI}_{ij}) \tag{15.12}$$

*Step* 6: Perform ANOVA for identifying the significant factors.

## ILLUSTRATION 15.6

The problem considered here is same as in Illustration 15.5. The S/N ratio values are calculated for appropriate quality characteristic using Eqs. (15.8) and (15.9) and are tabulated in Table 15.16.

**TABLE 15.16** Experimental results and S/N ratio for Illustration 15.6

| *Exp. no.* | *Control factors* | | | | *Tensile strength* (TS) | | *Surface roughness* (SR) | | *S/N ratio values* | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *A* | *B* | *C* | *Error* | 1 | 2 | 1 | 2 | TS | SR |
| 1 | 1 | 1 | 1 | 1 | 1075 | 1077 | 0.3892 | 0.3896 | 60.6362 | 8.19208 |
| 2 | 1 | 2 | 2 | 2 | 1044 | 1042 | 0.3397 | 0.3399 | 60.3657 | 9.37553 |
| 3 | 1 | 3 | 3 | 3 | 1062 | 1062 | 0.6127 | 0.6127 | 60.5225 | 4.25504 |
| 4 | 2 | 1 | 2 | 3 | 1036 | 1032 | 0.964 | 0.962 | 60.2904 | 0.32747 |
| 5 | 2 | 2 | 3 | 1 | 988 | 990 | 0.4511 | 0.4515 | 59.9039 | 6.91069 |
| 6 | 2 | 3 | 1 | 2 | 985 | 983 | 0.3736 | 0.3736 | 59.8599 | 8.55186 |
| 7 | 3 | 1 | 3 | 2 | 926 | 926 | 1.2712 | 1.271 | 59.3322 | –2.0836 |
| 8 | 3 | 2 | 1 | 3 | 968 | 966 | 1.291 | 1.2908 | 59.7085 | –2.2179 |
| 9 | 3 | 3 | 2 | 1 | 957 | 959 | 0.1557 | 0.1557 | 59.6273 | 16.1542 |

The normalized S/N ratio values are computed using Eqs. (15.3) and (15.4) and the same is given in Table 15.17.

**TABLE 15.17** Normalized S/N ratio for Illustration 15.6

| *Exp.* | *S/N ratio values* | | *Normalized S/N ratio values* | |
|---|---|---|---|---|
| *no.* | TS | SR | TS | SR |
| 1 | 60.63624543 | 8.192081 | 1 | 0.433383 |
| 2 | 60.36568617 | 9.375533 | 0.79252 | 0.368967 |
| 3 | 60.52249033 | 4.255042 | 0.912766 | 0.647678 |
| 4 | 60.29041078 | 0.327474 | 0.734795 | 0.861457 |
| 5 | 59.90392583 | 6.910693 | 0.438416 | 0.50313 |
| 6 | 59.85990197 | 8.551863 | 0.404656 | 0.4138 |
| 7 | 59.33221973 | –2.08359 | 0 | 0.992692 |
| 8 | 59.70852948 | –2.21785 | 0.288575 | 1 |
| 9 | 59.62731018 | 16.15423 | 0.226292 | 0 |

From the data in Table 15.17, factor analysis is performed based on principal component method. Here the components that are having eigenvalue more than one are selected for analysis. The component values obtained after performing factor analysis based on PCM are given in Table 15.18.

**TABLE 15.18** Component matrix for Illustration 15.6

| *Responses* | *Comp.* 1 |
|---|---|
| TS | 0.766 |
| SR | –0.766 |
| Eigen values | 1.174 |

Using the values from Table 15.18, the following MRPI equation is obtained,

$$MRPI_{i1} = 0.766\ Z_{i1} - 0.766\ Z_{i2}$$

where $Z_{i1}$, and $Z_{i2}$ represent the normalized S/N ratio values for the responses TS and SR at $i$th trial respectively. The MRPI values are computed and are tabulated in Table 15.19.

**TABLE 15.19** Normalized S/N ratios and MRPI values

| *Exp. no.* | *Normalized S/N ratio values* | | *MRPI* |
|---|---|---|---|
| | TS | SR | |
| 1 | 1 | 0.433383 | 0.434029 |
| 2 | 0.79252 | 0.368967 | 0.324441 |
| 3 | 0.912766 | 0.647678 | 0.203058 |
| 4 | 0.734795 | 0.861457 | –0.09702 |
| 5 | 0.438416 | 0.50313 | –0.04957 |
| 6 | 0.404656 | 0.4138 | –0.007 |
| 7 | 0 | 0.992692 | –0.7604 |
| 8 | 0.288575 | 1 | –0.54495 |
| 9 | 0.226292 | 0 | 0.17334 |

Table 15.20 summarizes the main effects on MRPI and Figure 15.1 shows plot of factor effects. The controllable factors on MRPI value in the order of significance are $A$, $B$, and $C$. The larger MRPI value implies the better quality. Consequently, the optimal condition is set as $A_1$, $B_3$ and $C_2$.

**TABLE 15.20** Main effects on MRPI

| *Factors* | 1 | 2 | 3 | *Max–Min* |
|---|---|---|---|---|
| Inlet temperature ($A$) | **0.320509** | –0.0512 | –0.37734 | 0.697847 |
| Injection time ($B$) | –0.14113 | –0.09003 | **0.123131** | 0.264263 |
| Injection pressure ($C$) | –0.03931 | **0.133586** | –0.2023 | 0.335891 |

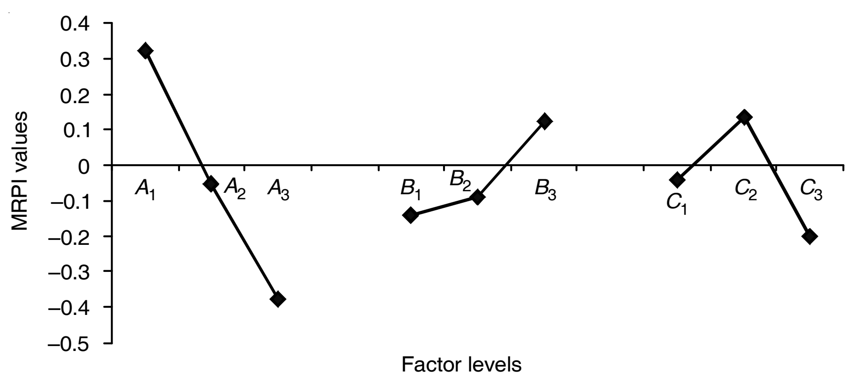



**FIGURE 15.1** Factor effects on MRPI.

The ANOVA on MRPI values is carried out and results are given in Table 15.21. The result of the pooled ANOVA shows that inlet temperature is the significant factor with a contribution of 58.97%.

**TABLE 15.21** Results of the pooled ANOVA

| *Factors* | *Sum of squares* | *Degrees of freedom* | *Mean squares* | *F* | *% contribution* |
|---|---|---|---|---|---|
| *A* | 0.731524 | 2 | 0.365762 | 4.308813 | 58.977 |
| *C* | 0.169283 | 2 | 0.084641 | 0.997108 | 13.64794 |
| Error | 0.339548 | 4 | 0.084887 | | 27.37506 |
| Total | 1.240355 | 8 | | | |

## 15.7 GENETIC ALGORITHM

Genetic Algorithms (GAs) are search heuristics used to solve global optimization problems in complex search spaces. GAs are intelligent random search strategies which have been used successfully to find near optimal solutions to many complex problems. GA has been applied successfully for solving problems combinatorial Optimization problems such as Travelling salesman problem, scheduling, vehicle routing problem, bin packing problem, etc., (Yenlay 2001). Jeyapaul et al. (2005) have applied GA to solve multi-response Taguchi problems. The detailed discussion and applications can be found in Goldberg (1989 and 1998).

The steps of GA to solve multi-response problems in Taguchi method is given below. The objective of this GA is to find the optimal weights so as to maximize the normalized S/N ratio. Here the weights are considered as genes and the sum of the weights should be equal to one.

*Step* 1: Transform the original data from the Taguchi experiment into S/N ratios for each type of response using the appropriate formula depending on the type of quality characteristic (Section 13.3).

*Step* 2: Normalize the S/N ratios as in the case of Illustration 15.5.

*Step* 3: The normalized S/N ratio values are considered for calculating the fitness value. The GA approach adopted in the robust design procedure is explained as follows:

- *Step* 3.1 *Initialization*: Randomly generate an initial population of *Ps* strings, where *Ps* is the population size.
- *Step* 3.2 *Evaluation*: Calculate the value of the objective function for each solution. Then transform the value of the objective function for each solution to the value of the fitness function for each string.
- *Step* 3.3 *Selection*: Select a pre-specified number of pairs of strings from the current population according to the selection methodology specified.
- *Step* 3.4 *Crossover*: Apply the pre-specified crossover operator to each of the selected pairs in Step 3.3 to generate *Ps* strings with the pre-specified crossover probability *Pc*.

*Step* 3.5 *Mutation*: Apply the pre-specified mutation operator to each of the generated strings with the pre-specified mutation probability *Pm*.

*Step* 3.6 *Termination test*: If a pre-specified stopping condition is satisfied, stop this algorithm. Otherwise, return to Step 3.2. The GA approach is depicted in Figure 15.2.

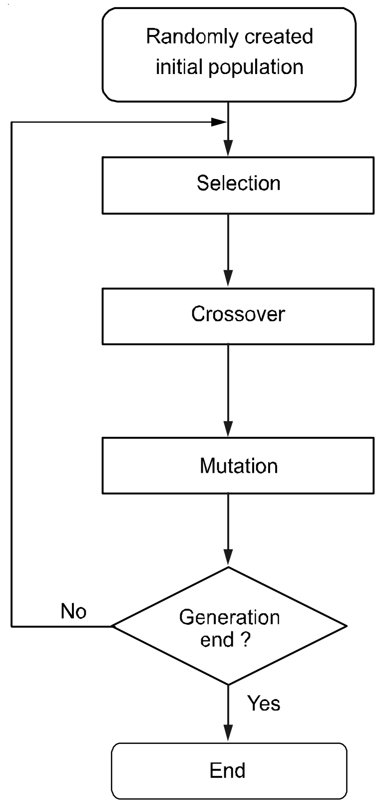



**FIGURE 15.2** Schematic representation of GA.

*Step* 4: Calculate weighted S/N ratio value using the weights ($w_1$, $w_2$, …,$w_j$) obtained from GA

$$\text{WSN}_i = W_1 Z_{11} + W_2 Z_{12} + \cdots + W_j Z_{ij} \quad (15.13)$$

*Step* 5: Determine the optimal level combination for the factors.

Maximization of weighted S/N ratio results in better product quality, therefore, on the basis of weighted S/N ratio, the factor effect is estimated and the optimal level for each controllable factor is determined. For Example, to estimate the effect of factor *i*, calculate the average of weighted S/N ratio values (WSN) for each level *j*, denoted as $\text{WSN}_{ij}$, then the effect $E_i$, is defined as:

$$E_i = \max(\text{WSN}_{ij}) - \min(\text{WSN}_{ij}) \quad (15.14)$$

If the factor $i$ is controllable, the best level $j^*$, is determined by

$$j^* = \max_j(\text{WSN}_{ij}) \tag{15.15}$$

*Step* 6: Perform ANOVA for identifying the significant factors.

### *The implementation of genetic algorithm*

The implementation of genetic algorithm is explained as follows:

***Initial population:*** In this algorithm, initialization that is initial population generation is often carried out randomly. The objective of this algorithm is to find the optimal weights to the responses to maximize the normalized S/N ratio. The number of genes in a chromosome is equal to the number of the responses. If we take five responses, the chromosome will be like [0.21 0.30 0.05 0.34 0.10] and the sum of the genes should be one. Here the weights are considered as the genes, initial populations of 20 chromosomes are generated subject to a feasibility condition, i.e., the sum of weights should equal one.

***Selection:*** Selection in evolutionary algorithms is defined by selection probabilities or rank which is calculated using the fitness value for each individual within a population. This selection probability can depend on the actual fitness values and hence they change between generations or they can depend on the rank of the fitness values only, which results in fixed values for all generations.

***Evaluation:*** The evaluation is finding the total weighted S/N ratio for the generated population. The objective function value for the problem is given

$$f(x) = \sum_{j=1}^{k} \sum_{i=1}^{n} W_j Z_{ij} \tag{15.16}$$

where

$f(x)$ = total weighted S/N (WSN) ratio to be maximized,
$W_j$ = weights for each response,
$Z_{ij}$ = normalized S/N ratio values,
$n$ = number of experiments under each response, and
$k$ = number of responses.

***Crossover:*** This is a genetic operator that combines two parent chromosomes to produce a new child chromosome. Here we use the single point crossover which is the most basic crossover operator where a single location is chosen and the offspring gets all the genes from the first parent up to that point and all the genes from the second parent after that point. Consider the following two parents which have been selected for crossover. The dotted line indicates the single crossover point.

***For Example:***

Parent 1 — Feasible condition: 1

| | | |
|---|---|---|
| 0.0658 | 0.8125 | 0.1217 |

Parent 2 — Feasible condition: 1

| | | |
|---|---|---|
| 0.2635 | 0.6692 | 0.0673 |

Offspring 1 — Feasible condition: 0.9456

| | | |
|---|---|---|
| 0.0658 | 0.8125 | 0.0673 |

Offspring 2 — Feasible condition: 1.0544

| | | |
|---|---|---|
| 0.2635 | 0.6692 | 0.1217 |

***Mutation:*** This allows new genetic patterns to be introduced that were not already contained in the population. The mutation can be thought of as natural experiments. These experiments introduce a new, somewhat random, sequence of genes into a chromosome.

***For Example:***
*Before mutation:*

Offspring 1 — Feasible condition: 1

| | | |
|---|---|---|
| 0.07485 | 0.83992 | 0.08543 |

*After mutation:*

Offspring 1 — Feasible condition: 1

| | | |
|---|---|---|
| 0.06886 | 0.83992 | 0.09142 |

### *Termination condition*

When the iteration meets the 10,000 number of iteration, the program comes to stopping condition.

The problem considered here is same as in Illustration 15.5. The S/N ratio values are calculated for appropriate quality characteristic using Eqs. (15.8) and (15.9) and are tabulated in Table 5.16. The normalized S/N ratio values are computed using Eqs. (15.3) and (15.4) and the same is given in Table 15.17.

By applying GA procedure to the normalized S/N ratios, the following result is obtained as optimal weights corresponding to each response. The optimal weights are [0.003256, 0.996735]. The WSN equation is:

$$\text{WSN}_{i1} = 0.003256\ Z_{i1} + 0.996735\ Z_{i2}$$

where $Z_{i1}$ and $Z_{i2}$, represent the normalized S/N ratio values for the responses TS and SR at *i*th trial respectively. The WSN values are computed and are listed in the last column of Table 15.22. Finally, the WSN values are considered for optimizing the multi-response parameter design problem.

**TABLE 15.22** Normalized S/N ratios and WSN values

| *Exp. no* | *Normalized S/N ratio values* | | WSN |
|---|---|---|---|
| | TS | SR | |
| 1 | 1 | 0.433383 | 0.435233 |
| 2 | 0.79252 | 0.368967 | 0.37035 |
| 3 | 0.912766 | 0.647678 | 0.648543 |
| 4 | 0.734795 | 0.861457 | 0.861043 |
| 5 | 0.438416 | 0.50313 | 0.502918 |
| 6 | 0.404656 | 0.4138 | 0.41377 |
| 7 | 0 | 0.992692 | 0.989451 |
| 8 | 0.288575 | 1 | 0.997677 |
| 9 | 0.226292 | 0 | 0.000739 |

Table 15.23 gives the main effects on WSN and Figure 15.3 shows the plot of the factor effects. The max-min column in Table 15.23 indicates the order of significance factors in affecting the process performance. The controllable factors on WSN value in the order of significance are *B*, *C* and *A*. The larger WSN value implies the better quality. So the optimal condition is set as $A_3$, $B_1$ and $C_3$. The results of the ANOVA (Table 15.24) gives that there is no significant factor.

**TABLE 15.23** Main effects on WSN values

| *Factors* | 1 | 2 | 3 | *Max–Min* |
|---|---|---|---|---|
| Inlet temperature (*A*) | 0.484709 | 0.592577 | **0.662622** | 0.177914 |
| Injection time (*B*) | **0.761909** | 0.623649 | 0.354351 | 0.407558 |
| Injection pressure (*C*) | 0.61556 | 0.410711 | **0.713638** | 0.302927 |

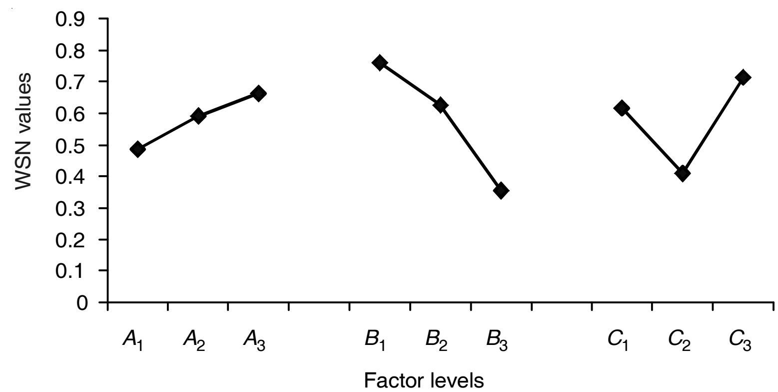



**FIGURE 15.3** Factor effects on WSN values.

**TABLE 15.24** Results of the pooled ANOVA

| *Factors* | *Sum of squares* | *Degrees of freedom* | *Mean squares* | *F* | *% Contribution* |
|---|---|---|---|---|---|
| *B* | 0.257741 | 2 | 0.128871 | 1.123722 | 29.97632 |
| *C* | 0.143347 | 2 | 0.071674 | 0.624977 | 16.67183 |
| Error | 0.458728 | 4 | 0.114682 | | 53.35185 |
| Total | 0.859816 | 8 | | | |

## PROBLEMS

**15.1** Six main factors *A*, *B*, *C*, *D*, *E* and *F* and one interaction *AB* has been studied and the following data have been obtained (Table 15.25). Note that there are two responses. Response 1 is smaller—the better type of quality characteristic and Response 2 is larger—the better type of characteristic.

**TABLE 15.25** Data for Problem 15.1

| *Trial* | *Factors/Columns* | | | | | | | *Response* 1 | | *Response* 2 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *no.* | *A* 1 | *B* 2 | *AB* 3 | *C* 4 | *D* 5 | *E* 6 | *F* 7 | 1 | 2 | 1 | 2 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4.15 | 3.8 | 12.95 | 13.55 |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 4.13 | 3.33 | 13.66 | 10.70 |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 3.15 | 2.02 | 13.29 | 10.39 |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2.99 | 2.64 | 13.60 | 14.20 |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 4.22 | 4.87 | 19.70 | 18.00 |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 5.74 | 6.53 | 20.11 | 19.41 |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 4.72 | 5.35 | 22.58 | 21.88 |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 3.27 | 4.07 | 13.27 | 11.57 |

Analyse the data by an appropriate weighting technique. And determine the optimal levels for the factors.

**15.2** Apply grey relational analysis to Problem 15.1 and determine optimal levels for the factors.

**15.3** Consider Problem 15.1 and apply Data Envelopment Analysis-based Ranking Method to determine optimal levels to the factors.

**15.4** Consider data in Problem 15.1. Assume that Response 1 is nominal—the best type and Response 2 is larger—the better type and apply grey relational analysis to obtain optimal levels for the factors.

**15.5** Consider the data given in $L_{18}$ OA (Table 15.26). The objective is to maximize the strength and minimize the roughness. Apply grey relational analysis and determine optimal levels for the factors.

**TABLE 15.26** Data for Problem 15.5

| *Trial* | *Factors/Columns* | | | | | | | | *Responses* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *no.* | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* | *Strength* | | *Roughness* | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 1 | 2 | 1 | 2 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 76.6 | 77.0 | 11.1 | 11.2 |
| 2 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 75.8 | 76.2 | 10.7 | 10.7 |
| 3 | 1 | 1 | 3 | 3 | 3 | 3 | 3 | 3 | 76.4 | 76.6 | 10.4 | 11.0 |
| 4 | 1 | 2 | 1 | 1 | 2 | 2 | 3 | 3 | 76.0 | 78.0 | 12.5 | 12.5 |
| 5 | 1 | 2 | 2 | 2 | 3 | 3 | 1 | 1 | 72.8 | 73.2 | 10.9 | 11.1 |
| 6 | 1 | 2 | 3 | 3 | 1 | 1 | 2 | 2 | 74.6 | 75.0 | 11.8 | 11.9 |
| 7 | 1 | 3 | 1 | 2 | 1 | 3 | 2 | 3 | 75.0 | 75.6 | 12.2 | 12.4 |
| 8 | 1 | 3 | 2 | 3 | 2 | 1 | 3 | 1 | 75.6 | 75.8 | 13.5 | 13.5 |
| 9 | 1 | 3 | 3 | 1 | 3 | 2 | 1 | 2 | 75.7 | 76.1 | 12.6 | 13.0 |
| 10 | 2 | 1 | 1 | 3 | 3 | 2 | 2 | 1 | 78.9 | 79.1 | 11.4 | 11.6 |
| 11 | 2 | 1 | 2 | 1 | 1 | 3 | 3 | 2 | 75.0 | 77.0 | 11.8 | 11.6 |
| 12 | 2 | 1 | 3 | 2 | 2 | 1 | 1 | 3 | 72.8 | 73.2 | 10.8 | 11.0 |
| 13 | 2 | 2 | 1 | 2 | 3 | 1 | 3 | 2 | 76.0 | 78.0 | 11.3 | 11.7 |
| 14 | 2 | 2 | 2 | 3 | 1 | 2 | 1 | 3 | 71.9 | 72.1 | 10.7 | 10.7 |
| 15 | 2 | 2 | 3 | 1 | 2 | 3 | 2 | 1 | 69.8 | 70.2 | 12.5 | 12.5 |
| 16 | 2 | 3 | 1 | 3 | 2 | 3 | 1 | 2 | 75.0 | 75.0 | 11.6 | 11.8 |
| 17 | 2 | 3 | 2 | 1 | 3 | 1 | 2 | 3 | 78.9 | 79.1 | 13.7 | 13.7 |
| 18 | 2 | 3 | 3 | 2 | 1 | 2 | 3 | 1 | 78.0 | 78.0 | 12.5 | 12.9 |

**15.6** Consider data in Problem 15.5. Apply Data Envelopment Analysis-based Ranking Approach to determine the optimal levels for the factors.

# CHAPTER 16

# Case Studies

## 16.1 MAXIMIZATION OF FOOD COLOR EXTRACT FROM A SUPER CRITICAL FLUID EXTRACTION PROCESS

This case study pertains to a company manufacturing natural colours, flavours and oleoresins from botanicals. Annato seed is one of the inputs from which food colour is extracted using Super Critical Fluid Extraction process (SCFE). SCFE is a non-chemical separation process operating in a closed system with liquid carbon dioxide gas at high hydraulic pressure and normal temperatures. Three important issues related to the natural colour extraction process are the yield, the colour value (quality or grade) and percentage of bio-molecules in the yield. Since the yield (output or extract) from SCFE is directly used in the colour formulation, any increase in the yield would benefit the company. The present level of average yield is 3.8% of input material. From the initial discussion with the concerned executives, it is learnt that the values for the process parameters have been arrived at by experience and judgement. Hence it was decided to optimize the process parameters using Taguchi Method. So, the objective of this study is to maximize the yield through parameter design.

### *Identification of factors*

The factors that could affect the process yield have been identified through brainstorming are: (i) Raw material type, (ii) Pressure, (iii) Raw material size, (iv) Time, (v) Flow rate, (vi) Oven temperature, (vii) Valve temperature, (viii) Carbon dioxide gas purity, and (ix) Chiller temperature.

The company uses only high purity carbon dioxide gas and hence this factor may not have any effect on the yield. The purpose of chiller temperature (–10° C to +10°C) is to keep carbon dioxide gas in liquid state and thus this may also not affect the yield. Hence the first seven factors listed above have been considered in this study.

***Levels of the factors:*** The present level settings have been fixed by experience. After discussing with the technical and production people concerned with the process, the two levels for each factor proposed to be studied are selected. The factors and their levels are given in Table 16.1. Due to confidentiality, the levels for the factors *A*, *B* and *C* are coded and given.

In addition to the main factors, the interactions *BE*, *BF* and *EG* are suspected to have an effect on the yield.

**TABLE 16.1** Process factors and their levels

| *Factors* | *Operating range* | *Present setting* | *Proposed setting* | |
|---|---|---|---|---|
| | | | *Level* 1 | *Level* 2 |
| Seed type (*A*) | AP and TN | TN | Type 1 | Type 2 |
| Pressure (*B*) | 350–700 bar | $X_1$ | $X_2$ | $X_3$ |
| Particle size (*C*) | $M_1$, $M_2$, $M_3$ | $M_2$ | $M_1$ | $M_3$ |
| Time (*D*) | 0.5–1.0 hr | 0.5 hr | 0.75 hr | 1.0 hr |
| Flow rate (*E*) | 1.5–3 units | 2 units | 2.5 | 3.0 |
| Oven temp. (*F*) | 70°C–110°C | 80°C | 70°C | 100°C |
| Valve temp. (*G*) | 90°C–130°C | 90°C | 100°C | 120°C |

***Response variable:*** The SCFE process is being used for colour extraction from annatto seeds. The percentage yield, the colour value and the percentage of bio-molecules present could be used as response variables. However, in this study percentage yield has been selected as the response variable as per the objective. It is computed as follows:

$$\text{Percentage yield} = \left(\frac{X - Y}{W}\right) \times 100$$

where

$X$ = final weight of the collection bottle
$Y$ = initial weight of the collection bottle
$W$ = total weight of raw materials

Weight of raw material has been maintained at a constant value in all the experiments.

***Design of the experiment and data collection:*** There are seven main factors and three interactions to be studied and have 10 degrees of freedom between them. Hence $L_{16}$ OA has been selected. The assignment of factors and interactions are given in Table 16.2. Triangular table has been used to determine the assignment of the main factors and interactions to the OA. Sixteen trials have been conducted randomly each with two replications. Simple repetition has been used to replicate the experiment. The layout of the experiment with the response (yield in percentage) is tabulated in Table 16.2.

*Data analysis:* The response totals is provided in Table 16.3.

Grand total = 119.19
Total number of observations = 32

$$\text{Correction factor (CF)} = \frac{(119.19)^2}{32} = 443.9455$$

$$\begin{aligned} SS_{\text{Total}} &= (4.47)^2 + (3.93)^2 + (6.19)^2 + \cdots + (2.00)^2 + (3.70)^2 - \text{CF} \\ &= 487.6773 - 443.9455 \\ &= 43.7318 \end{aligned}$$

**TABLE 16.2** Layout of the experiment and the data for Case 16.1

| *Trial. no.* | *B* 1 | *E* 2 | *BE* 3 | *G* 4 | *EG* 6 | *A* 8 | *C* 10 | *D* 12 | *F* 14 | *BF* 15 | Response $Y_1$ | $Y_2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4.47 | 4.47 |
| 2 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 3.93 | 4.03 |
| 3 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 6.19 | 5.97 |
| 4 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 3.43 | 3.20 |
| 5 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 2.33 | 2.03 |
| 6 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 3.83 | 4.26 |
| 7 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 3.67 | 2.13 |
| 8 | 1 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 4.30 | 4.33 |
| 9 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 4.53 | 4.06 |
| 10 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 4.67 | 4.60 |
| 11 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 5.27 | 5.37 |
| 12 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2.10 | 2.63 |
| 13 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2.03 | 1.70 |
| 14 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 2.93 | 3.80 |
| 15 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 3.03 | 2.00 |
| 16 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 4.20 | 3.70 |

**TABLE 16.3** Response totals for Case 16.1

| *Factors* | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *BE* | *EG* | *BF* |
|---|---|---|---|---|---|---|---|---|---|---|
| Level 1 | 59.25 | 62.57 | 71.68 | 53.51 | 68.92 | 54.54 | 57.67 | 59.08 | 62.12 | 61.00 |
| Level 2 | 59.94 | 56.62 | 47.51 | 65.68 | 50.27 | 64.65 | 61.52 | 60.11 | 57.07 | 59.19 |

$$
\begin{aligned}
SS_A &= \frac{A_1^2}{n_{A1}} + \frac{A_2^2}{n_{A2}} - \text{CF} \\
&= \frac{(59.25)^2}{16} + \frac{(59.94)^2}{16} - \text{CF} \\
&= 443.9604 - 443.9455 \\
&= 0.0149
\end{aligned}
$$

Similarly, the sum of squares of all effects are computed and tabulated in Table 16.4.

From the initial ANOVA (Table 16.4), it can be seen that the main effects *B*, *C*, *D*, *E*, and *F* are significant. Also, the interaction *EG* seems to have some influence ($F = 3.92$). The other effects are pooled into the error term and the final ANOVA is given in Table 16.5. From the final ANOVA (Table 16.5), it is found that only the main effects *B*, *C*, *D*, *E* and *F* are significant.

**TABLE 16.4** ANOVA (initial) for Case 16.1

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ | *C*(%) |
|---|---|---|---|---|---|
| *A* | 0.0149 | 1 | 0.0149 | 0.07 | 0.03 |
| *B* | 1.1063 | 1 | 1.1063 | 5.44* | 2.53 |
| *C* | 18.2559 | 1 | 18.2559 | 89.84* | 41.75 |
| *D* | 4.6284 | 1 | 4.6284 | 22.79* | 10.58 |
| *E* | 10.8695 | 1 | 10.8695 | 53.49* | 24.85 |
| *F* | 3.1941 | 1 | 3.1941 | 15.72* | 7.30 |
| *G* | 0.4632 | 1 | 0.4632 | 2.28 | 1.06 |
| *BE* | 0.0332 | 1 | 0.0332 | 0.16 | 0.08 |
| *EG* | 0.7970 | 1 | 0.7970 | 3.92 | 1.82 |
| *BF* | 0.1024 | 1 | 0.1024 | 0.50 | 0.23 |
| Error | 4.2669 | 21 | 0.2032 | | 9.77 |
| Total | 43.7318 | 31 | | | 100.00 |

$F_{0.05,1,21} = 4.32$; *Significant at 5% level.

**TABLE 16.5** ANOVA (final) for Case 16.1

| *Source of variation* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ | *C*(%) |
|---|---|---|---|---|---|
| *B* | 1.1063 | 1 | 1.1063 | 5.67 | 2.54 |
| *C* | 18.2559 | 1 | 18.2559 | 93.52 | 41.75 |
| *D* | 4.6284 | 1 | 4.6284 | 23.71 | 10.58 |
| *E* | 10.8695 | 1 | 10.8695 | 55.68 | 24.85 |
| *F* | 3.1941 | 1 | 3.1941 | 16.36 | 7.30 |
| *EG* | 0.7970 | 1 | 0.7970 | 4.08 | 1.82 |
| Pooled error | 4.8806 | 25 | 0.1952 | | 11.16 |
| Total | 43.7318 | 31 | | | 100.00 |

Since $F_{0.05,1,25} = 4.24$, the interaction *EG* is not significant at 5% level of significance. And all the main effects given in Table 16.5 are significant. However, the interaction *EG* seems to have some influence ($F = 4.08$). Hence this interaction has been considered while determining the optimal levels for the factors.

***Optimal levels:*** The optimal levels are selected based on the average response of the significant factors given in Table 16.6.

**TABLE 16.6** Average response for Case 16.1

| *Factors* | *B* | *C* | *D* | *E* | *F* |
|---|---|---|---|---|---|
| Level 1 | **3.91** | **4.48** | 3.34 | **4.31** | 3.41 |
| Level 2 | 3.54 | 2.97 | **4.11** | 3.14 | **4.04** |

For the interaction (EG), the response totals for the four combinations and their average is given in Table 16.7.

**TABLE 16.7** Response totals for Case 16.1

| *Interaction* | $E_1G_1$ | $E_1G_2$ | $E_2G_1$ | $E_2G_2$ |
|---|---|---|---|---|
| Response total | 34.76 | 34.16 | 22.91 | 27.36 |
| Average response | **4.35** | 4.27 | 2.86 | 3.42 |

Since the objective is to maximize the yield, the optimal condition is $B_1C_1D_2E_1F_2G_1$. The optimal levels recommended are given in Table 16.8.

**TABLE 16.8** Optimal levels recommended for Case 16.1

| *Factors* | *Level recommended* |
|---|---|
| Seed type ($A$) | Type 1 or Type 2 |
| Pressure ($B$) | Level 1 = $X_2$ |
| Particle size ($C$) | Level 1 = $M_1$ |
| Time ($D$) | Level 2 = 1.0 hr |
| Flow rate ($E$) | Level 1 = 2.5 units |
| Oven temperature ($F$) | Level 2 = 100°C |
| Valve temperature ($G$) | Level 1 = 100°C |

The factors $C$, $D$, $E$ and $F$ together contribute about 85% of the total variation. So, for predicting the optimal yield, these four are considered.

$$\overline{Y} = \frac{119.19}{32} = 3.72$$

$$\overline{C}_1 = 4.48;\ \overline{D}_2 = 4.11;\ \overline{E}_1 = 4.31;\ \overline{F}_2 = 4.04$$

$$\mu_{\text{Pred}} = \overline{Y} + (\overline{C}_1 - \overline{Y}) + (\overline{D}_2 - \overline{Y}) + (\overline{E}_1 - \overline{Y}) + (\overline{F}_2 - \overline{Y}) \tag{16.1}$$

$$= \overline{C}_1 + \overline{D}_2 + \overline{E}_1 + \overline{F}_2 - 3\overline{Y}$$

$$= (4.48 + 4.11 + 4.31 + 4.04) - 3(3.72)$$

$$= 16.94 - 11.16$$

$$= 5.78$$

***Confirmation experiment:*** A confirmation experiment at the recommended levels has been run. From three replications with Type 1 seed and three replications from Type 2 seed, the average yield has been found to be 5.82%, which is almost same as $\mu_{\text{Pred}}$.

***Conclusion:*** The present yield of the colour extraction from the SCFE process is 3.8%. After experimentation, the levels have been changed and with these levels the yield has been increased to about 5.8%. Thus, the yield has been increased by about 52% from the present level.

***Source:*** Krishnaiah, K., and Rajagopal, K., Maximization of food colour extract from a super critical fluid extraction process—a case study, *Journal of Indian Institution of Industrial Engineering*, vol. XXX, no. 1, January 2001.

## 16.2 AUTOMOTIVE DISC PAD MANUFACTURING

This case has been conducted in a company manufacturing automotive disc pads used in brakes. The problem was the occurrence of relatively high number of rejections in the disc pad module. The average rejection rate in the disc pad was observed to be about 1.2%. Though it might seem to be less, it is necessary to further reduce the rejections in today's world of 'zero ppm'.

In order to know which type of defect contribute more to the defectives, Pareto analysis had been performed. The Pareto chart is shown in Figure 16.1.

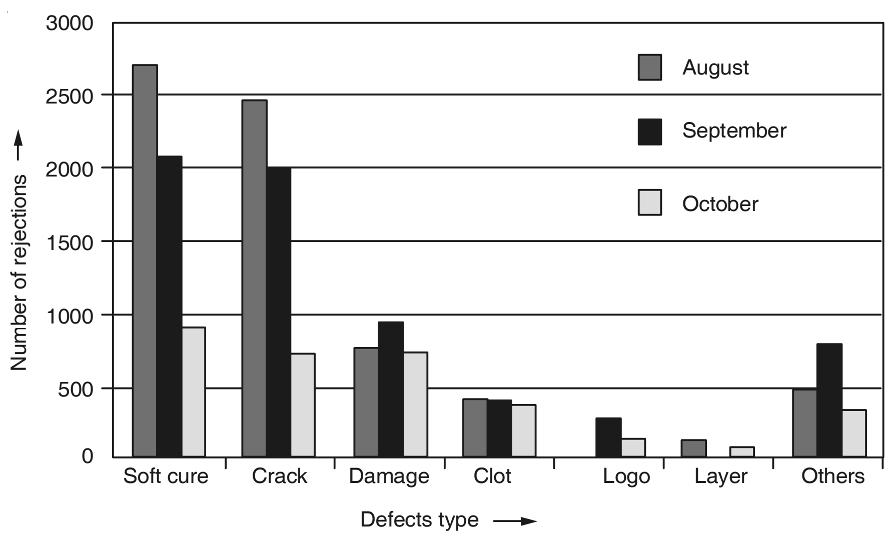
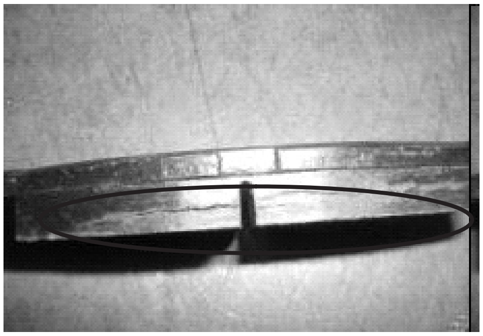


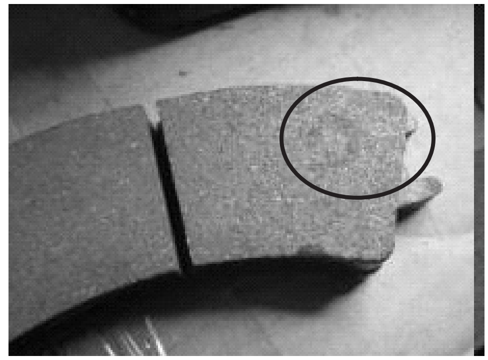

**FIGURE 16.1** Pareto analysis of defects in disc pad.

From Figure 16.1, it is evident that the major defects affecting the disc pad are the crack and the soft cure shown in Figure 16.2.

***Objective of the study:*** The objective of the study was to reduce the rejections in the disc pad module of the company, thereby improving the productivity of the disc pad module. In order to reduce the rejections, the two major defects namely, crack and soft cure are considered.



**Crack** (An imperfection above/below the surface of the cured pad)



**Soft Cure** (Presence of soft areas in the cured pad)

FIGURE 16.2 Major defects of disc pad.

The cause and effect diagram for the two types of defects is shown in Figures 16.3 and 16.4.

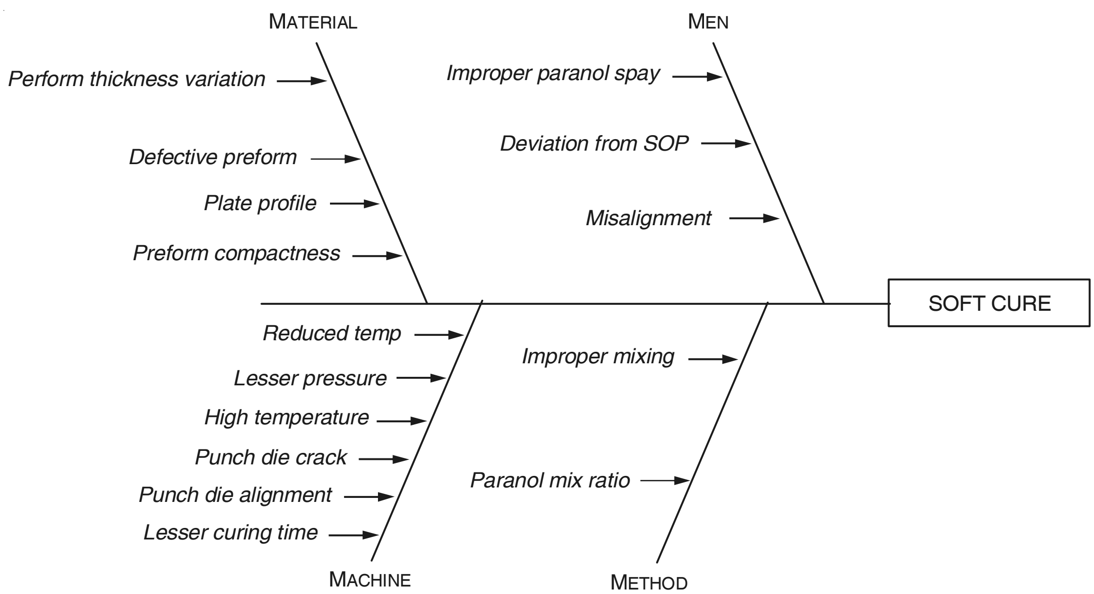



FIGURE 16.3 Cause and effect diagram for the defect soft cure.

***Identification of factors and their levels:*** After analysing the cause and effect diagrams, the team of the quality and manufacturing engineers and the experimenter decided to have the following factors for study. The levels were also fixed in consultation with the engineers. Factors and their current levels, and proposed levels for the factors are given in Table 16.9 and 16.10 respectively.

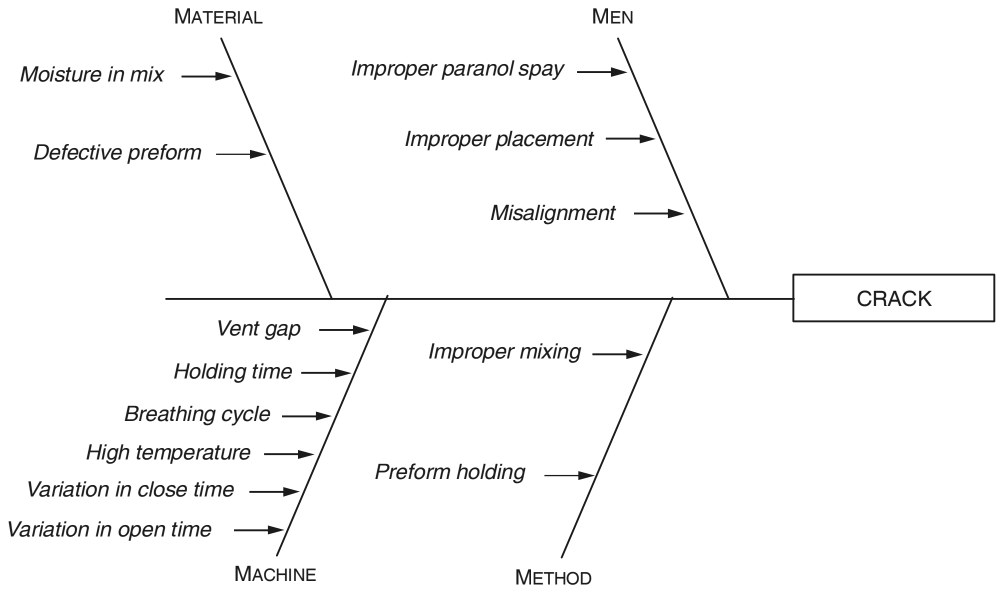



**FIGURE 16.4** Cause and effect diagram for the defect crack.

**TABLE 16.9** Factors and their current levels

| *Factors* | *Operating ranges* | *Current levels* |
|---|---|---|
| Temperature | 143°C–156°C | 150°C |
| Pressure | 165–175 kgf/cm² | 175 kgf/cm² |
| Close time | 3.5 s–4.5 s | 4.1 s |
| Open time | 3.5 s–4.5 s | 3.65 s |
| No. of vents | 2–4 | 3 |
| Paranol mix ratio | 1:10–1:20 | 1:10 |

**TABLE 16.10** Proposed levels for the factors

| *Code* | *Factors* | | *Levels* | |
|---|---|---|---|---|
| A | Temperature | 143°C–149°C | — | 150°C–156°C |
| H | Pressure | 165 kgf/cm² | 170 kgf/cm² | 175 kgf/cm² |
| C | Close time | 3.5 s | 4.0 s | 4.5 s |
| D | Open time | 3.5 s | 4.0 s | 4.5 s |
| E | No. of vents | 2 | 3 | 4 |
| G | Paranol mix ratio | 1:10 | 1:15 | 1:20 |

*Design of the experiment:* Since one factor is at two levels and others are at three levels, $L_{18}$ OA has been selected and the factors are randomly assigned to the columns (Table 16.11). The vacant columns are assigned by $e$ indicating error.

**TABLE 16.11** Assignment of factors for Case 16.2

| *Trial no.* | *Columns/Factors* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| | $A$ | $e_2$ | $C$ | $D$ | $E$ | $e_6$ | $G$ | $H$ |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 1 | 1 | 3 | 3 | 3 | 3 | 3 | 3 |
| 4 | 1 | 2 | 1 | 1 | 2 | 2 | 3 | 3 |
| 5 | 1 | 2 | 2 | 2 | 3 | 3 | 1 | 1 |
| 6 | 1 | 2 | 3 | 3 | 1 | 1 | 2 | 2 |
| 7 | 1 | 3 | 1 | 2 | 1 | 3 | 2 | 3 |
| 8 | 1 | 3 | 2 | 3 | 2 | 1 | 3 | 1 |
| 9 | 1 | 3 | 3 | 1 | 3 | 2 | 1 | 2 |
| 10 | 2 | 1 | 1 | 3 | 3 | 2 | 2 | 1 |
| 11 | 2 | 1 | 2 | 1 | 1 | 3 | 3 | 2 |
| 12 | 2 | 1 | 3 | 2 | 2 | 1 | 1 | 3 |
| 13 | 2 | 2 | 1 | 2 | 3 | 1 | 3 | 2 |
| 14 | 2 | 2 | 2 | 3 | 1 | 2 | 1 | 3 |
| 15 | 2 | 2 | 3 | 1 | 2 | 3 | 2 | 1 |
| 16 | 2 | 3 | 1 | 3 | 2 | 3 | 1 | 2 |
| 17 | 2 | 3 | 2 | 1 | 3 | 1 | 2 | 3 |
| 18 | 2 | 3 | 3 | 2 | 1 | 2 | 3 | 1 |

*Data collection:* In each shift, they produce 360 disc pads. The occurrence of defectives also has been less. Hence each shift's production has been considered as one sample. That is from each experiment the output was 360 items. These 360 items have been subjected to 100% inspection and defectives were recorded (both types of defects included). The order of experiments was random. The data collected is given in Table 16.12.

The defectives are converted into fraction defective and transformed into S/N ratio using Eq. (13.4) which is written below. The S/N data is also given in Table 16.12.

$$\text{S/N ratio} = 10 \log\left(\frac{p}{1 - p}\right) \tag{16.2}$$

***Data analysis (Response Graph Method):*** The response (S/N) totals of all the factor effects and the average response is tabulated in Tables 16.13 and 16.14 respectively. The ranking of factor effects is given in Table 16.14.

**TABLE 16.12** Data for Case 16.2

| *Standard $L_{18}$ OA Trial no.* | *Defectives per 360 items* | *Fraction defective (p)* | *S/N ratio* |
|---|---|---|---|
| 1 | 4 | 0.0111 | –19.50 |
| 2 | 6 | 0.0167 | –17.70 |
| 3 | 0 | 0 | 0 |
| 4 | 2 | 0.0056 | –22.49 |
| 5 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 |
| 7 | 3 | 0.0083 | –20.77 |
| 8 | 0 | 0 | 0 |
| 9 | 1 | 0.0028 | –25.52 |
| 10 | 0 | 0 | 0 |
| 11 | 1 | 0.0028 | –25.52 |
| 12 | 1 | 0.0028 | –25.52 |
| 13 | 7 | 0.0194 | –17.04 |
| 14 | 0 | 0 | 0 |
| 15 | 5 | 0.0139 | –18.51 |
| 16 | 3 | 0.0083 | –20.77 |
| 17 | 4 | 0.0111 | –19.50 |
| 18 | 1 | 0.0028 | –25.52 |

**TABLE 16.13** Response totals for Case 16.2

| *Factors* | *A* | *C* | *D* | *E* | *G* | *H* | $e_2$ | $e_6$ |
|---|---|---|---|---|---|---|---|---|
| Level 1 | –105.98 | –100.57 | –131.04 | –91.31 | –91.31 | –63.53 | –88.24 | –81.56 |
| Level 2 | –152.38 | –62.72 | –106.55 | –104.99 | –76.48 | –106.55 | –58.04 | –91.23 |
| Level 3 | – | –95.07 | –20.7 | –62.03 | –90.57 | –88.28 | –112.08 | –85.57 |

**TABLE 16.14** Average response and ranking of factor effects for Case 16.2

| *Factors* | *A* | *C* | *D* | *E* | *G* | *H* |
|---|---|---|---|---|---|---|
| Level 1 | –11.78 | –16.76 | –21.84 | –15.22 | –15.22 | –10.59 |
| Level 2 | –16.93 | –10.45 | –17.76 | –17.50 | –12.75 | –17.76 |
| Level 3 | – | –15.85 | –3.46 | –10.34 | –15.10 | –14.71 |
| Difference | 5.15 | 6.31 | 18.38 | 7.16 | 2.47 | 7.17 |
| Rank | 5 | 4 | 1 | 3 | 6 | 2 |

***Optimum levels for the factors:*** The best levels for the factors are selected based on maximum S/N ratio (Table 16.14). Accordingly the best (optimum) levels for the factors in the order of increasing ranking are given in Table 16.15.

**TABLE 16.15** Factors and their levels for Case 16.2

| Factors | *D* | *H* | *E* | *C* | *A* | *G* |
|---|---|---|---|---|---|---|
| Level | 3 | 1 | 3 | 2 | 1 | 2 |

The optimal levels for the factors in terms of their original value are given in Table 16.16.

**TABLE 16.16** Optimum level of factors for Case 16.2

| *Factors* | *Optimal level* |
|---|---|
| Temperature (*A*) | 143°C–149°C |
| Close time (*B*) | 4.0 s |
| Open time (*C*) | 4.5 s |
| No. of vents (*D*) | 4 |
| Paranol mix. (*E*) | 1:15 |
| Pressure (*F*) | 165 kgf/cm² |

***Confirmation experiment:*** A confirmation experiment has been run using the proposed optimum parameter levels. The experiment has been run for three replicates (three shifts). From the confirmation experiment, the results obtained for three consecutive shifts were zero defectives. This result validates the proposed optimum levels for the factors.

***Source:*** Quality Improvement using Taguchi Method, Unpublished undergraduate project report supervised by Krishnaiah, K., (Professor), Department of Industrial Engineering, Anna University, 2007.

## 16.3 A STUDY ON THE EYE STRAIN OF VDT USERS

An experiment has been designed to investigate some of the factors affecting eye strain and suggest better levels for the factors so as to minimize the eye strain.

***Selection of factors and levels:*** There are several factors which affect the eye strain such as orientation of display screen, lighting, design of work station, nature of work, exposure time, reference material, environmental conditions, radiation and heat. The following factors and levels (Table 16.17) have been selected for study after discussing with eye specialists.

**TABLE 16.17** Factors and their proposed levels for Case 16.3

| *Factors* | *Proposed levels of study* | |
|---|---|---|
| | *Level* 1 | *Level* 2 |
| Viewing angle (*A*) | 0° (at eye level) | –20° (below eye level) |
| Viewing distance (*B*) | 40 cm | 70 cm |
| Exposure time (*C*) | 60 min | 120 min |
| Illumination (*D*) | 250 lux | 600 lux |

***Design of the experiment:*** Since the study involves four factors each at two levels, we have a $2^4$ full factorial design. That is, all the main factors and their interactions can be studied. Hence, $L_{16}$ OA has been selected for study. The assignment of factors and interactions is given in Table 16.18.

**TABLE 16.18** Assignment of factors and interactions for Case 16.3

| *Trial no.* | *Columns/Factor effects* | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| | $A$ | $B$ | $A$<br>$B$ | $A$<br>$B$<br>$C$ | $B$<br>$D$ | $A$<br>$B$<br>$D$ | $C$<br>$D$ | $C$ | $A$<br>$C$ | $B$<br>$C$ | $A$<br>$C$<br>$D$ | $D$ | $A$<br>$D$ | $B$<br>$C$<br>$D$ | $A$<br>$B$<br>$C$<br>$D$ |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 |
| 4 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 |
| 5 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 |
| 6 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 |
| 7 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 |
| 8 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 |
| 9 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| 10 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 |
| 11 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 |
| 12 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 |
| 13 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 |
| 14 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 |
| 15 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 |
| 16 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 |

***Selection of samples:*** The study has been carried out over a selected sample of subjects within the age group of 20–25 years having normal vision. They all had same level of experience.

***Response variable:*** The extent of eye strain has been measured by dynamic retinoscopy. It was measured by estimating the lag of accommodation.

***Data collection:*** The study has been carried out in a leading eye hospital in Chennai, India. Four replications have been obtained. The experimental results are given in Table 16.19.

***Data analysis:*** Data has been analysed using ANOVA. The initial ANOVA is given in Table 16.20.

**TABLE 16.19** Experimental results for Case 16.3

| *Standard $L_{16}$ OA Trial no.* | $R_1$ | $R_2$ | $R_3$ | $R_4$ | *Response total* |
|---|---|---|---|---|---|
| 1 | 1.50 | 1.50 | 1.50 | 1.50 | 6.00 |
| 2 | 1.75 | 1.75 | 1.50 | 1.75 | 6.75 |
| 3 | 1.50 | 1.25 | 1.25 | 1.50 | 5.50 |
| 4 | 2.00 | 2.00 | 2.00 | 2.00 | 8.00 |
| 5 | 0.65 | 0.40 | 0.65 | 0.65 | 2.35 |
| 6 | 0.65 | 0.90 | 0.65 | 0.90 | 3.10 |
| 7 | 0.40 | 0.65 | 0.40 | 0.65 | 2.10 |
| 8 | 0.90 | 0.90 | 0.90 | 0.65 | 3.35 |
| 9 | 1.25 | 1.50 | 1.50 | 1.50 | 5.75 |
| 10 | 1.50 | 1.50 | 1.75 | 1.50 | 6.25 |
| 11 | 1.75 | 1.50 | 1.50 | 1.50 | 6.25 |
| 12 | 2.00 | 2.00 | 2.00 | 2.00 | 8.00 |
| 13 | 0.90 | 0.65 | 0.65 | 0.65 | 2.85 |
| 14 | 0.65 | 0.65 | 0.90 | 0.65 | 2.85 |
| 15 | 0.40 | 0.65 | 0.40 | 0.40 | 1.85 |
| 16 | 0.90 | 0.90 | 0.90 | 0.90 | 3.60 |

Grand total = 74.55

**TABLE 16.20** ANOVA (initial) for Case 16.3

| *Source* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ |
|---|---|---|---|---|
| *A* | 0.0009 | 1 | 0.0009 | 0.06 |
| *B* | 14.4875 | 1 | 14.4875 | 999.13* |
| *AB* | 0.0009 | 1 | 0.0009 | 0.06 |
| *ABC* | 0.1181 | 1 | 0.1181 | 8.14* |
| *BD* | 0.0244 | 1 | 0.0244 | 1.68 |
| *ABD* | 0.1650 | 1 | 0.1650 | 11.37* |
| *CD* | 0.0478 | 1 | 0.0478 | 3.29 |
| *C* | 1.3369 | 1 | 1.3369 | 92.20* |
| *AC* | 0.0244 | 1 | 0.0244 | 1.68 |
| *BC* | 0.0478 | 1 | 0.0478 | 3.29 |
| *ACD* | 0.0087 | 1 | 0.0087 | 0.60 |
| *D* | 0.4306 | 1 | 0.4306 | 29.69* |
| *AD* | 0.0087 | 1 | 0.0087 | 0.60 |
| *BCD* | 0.0087 | 1 | 0.0087 | 0.60 |
| *ABCD* | 0.0478 | 1 | 0.0478 | 3.29 |
| Error | 0.7001 | 48 | 0.0145 | |
| Total | 17.4583 | 63 | | |

$F_{0.05,1,48} = 4.048$; *Significant at 5% level.

The insignificant effects are pooled into the error term so that approximately one-half of the effects are retained in the final ANOVA (Table 16.21).

**TABLE 16.21** ANOVA (final) for Case 16.3

| *Source* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ | *C(%)* |
|---|---|---|---|---|---|
| *B* | 14.4875 | 1 | 14.4875 | 1025.76* | 82.98 |
| *C* | 1.3369 | 1 | 1.3369 | 94.65* | 7.66 |
| *D* | 0.4306 | 1 | 0.4306 | 30.48* | 2.47 |
| *BC* | 0.0478 | 1 | 0.0478 | 3.38 | 0.27 |
| *CD* | 0.0478 | 1 | 0.0478 | 3.38 | 0.27 |
| *ABC* | 0.1181 | 1 | 0.1181 | 8.36* | 0.68 |
| *ABD* | 0.1650 | 1 | 0.1650 | 11.68* | 0.95 |
| *ABCD* | 0.0478 | 1 | 0.0478 | 3.38 | 0.27 |
| Pooled error | 0.7768 | 55 | 0.0141 | | 4.45 |
| Total | 17.4583 | 63 | | | 100.00 |

$F_{0.05,1,55} = 4.02$; *Significant at 5% level.

***Optimum levels for the factors:*** From Table 16.21, it is found that the main factors *B*, *C*, *D* and the three factor interactions *ABC* and *ABD* are significant. Hence, for determining the optimum levels for the factors the average response for all possible combinations of the three factor interactions *ABC* and *ABD* are evaluated and given in Table 16.24. Since the objective is to minimize the eye strain, the combinations $A_1B_2C_1$ and $A_2B_2D_2$ are selected (Table 16.24). We have to find the optimal levels for all the four factors *A*, *B*, *C* and *D*. Since *D* is not present in $A_1B_2C_1$ and *C* is not appearing in $A_2B_2D_2$, the following combinations are evaluated (Table 16.22).

**TABLE 16.22** Evaluation of ABCD interaction for Case 16.3

| *Combination* | *Average response* |
|---|---|
| $A_1B_2C_1D_1$ | 0.5250 |
| $A_1B_2C_1D_2$ | 0.5875 |
| $\boldsymbol{A_2B_2C_1D_2}$ | **0.4625** |
| $A_1B_2C_2D_2$ | 0.7125 |

For minimization of eye strain, the optimum condition is $\boldsymbol{A_2B_2C_1D_2}$. The optimal levels for the factors are given in Table 16.23.

**TABLE 16.23** Optimal levels of factor for Case 16.3

| *Factors* | *Level* |
|---|---|
| Viewing angle (*A*) | –20° |
| Viewing distance (*B*) | 70 cm |
| Exposure time (*C*) | 60 cm |
| Illumination (*D*) | High (600 lux) |

**TABLE 16.24** Evaluations of *ABC* and *ABD* Interactions for Case 16.3

| *ABC Interaction* | | *ABD Interaction* | |
|---|---|---|---|
| *Combination* | *Average response* | *Combination* | *Average response* |
| $A_1B_1C_1$ | 1.4218 | $A_1B_1D_1$ | 1.7500 |
| $A_1B_1C_2$ | 1.8434 | $A_1B_1D_2$ | 1.5312 |
| $\mathbf{A_1B_2C_1}$ | **0.5562** | $A_1B_2D_1$ | 0.7125 |
| $A_1B_2C_2$ | 0.8662 | $A_1B_2D_2$ | 0.6500 |
| $A_2B_1C_1$ | 1.5000 | $A_2B_1D_1$ | 1.7187 |
| $A_2B_1C_2$ | 1.7812 | $A_2B_1D_2$ | 1.5625 |
| $A_2B_2C_1$ | 0.5875 | $A_2B_2D_1$ | 0.8062 |
| $A_2B_2C_2$ | 0.8062 | $\mathbf{A_2B_2D_2}$ | **0.5875** |

***Confirmation experiment:*** This was conducted with the proposed optimal levels and five replications were made. The data from the confirmation experiment is given in Table 16.25.

**TABLE 16.25** Data for confirmation experiment for Case 16.3

| Trial no. | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Response | 0.65 | 0.40 | 0.40 | 0.65 | 0.65 |

At the optimal condition $A_2B_2C_1D_2$, the mean value ($\mu$) obtained is 0.55.

***Optimum predicted response:*** The grand average $(\bar{Y}) = \dfrac{74.55}{64} = 1.1648.$

The optimum predicted response at the optimum condition is

$$\begin{aligned}\mu_{\text{pred}} &= \bar{Y} + (A_2B_2C_1D_2 - \bar{Y}) \\ &= 1.1648 + (0.4625 - 1.1648) \\ &= 0.4625\end{aligned} \tag{16.3}$$

Confidence interval for confirmation experiment

$$\text{CI} = \sqrt{F_{\alpha,v1,v2} \times \text{MS}_e \times [(1/n_{\text{eff}}) + 1/r]} \tag{16.4}$$

$F_{0.05,1,55} = 4.02$

$$n_{\text{eff}} = N/(1 + \text{total df considered for prediction})$$

$$= \frac{64}{(1+1)} = 32$$

$r$ = number of replications in confirmation experiment = 5

$MS_e$ = Error mean square = 0.0141

$$\text{CI} = \sqrt{4.02 \times 0.0141 \times \left[\left(\frac{1}{32}\right) + \frac{1}{5}\right]} = 0.1145$$

The mean obtained from confirmation experiment ($\mu$) = 0.55

Therefore,

The confidence interval upper limit = ($\mu_{\text{Pred}}$ + CI) = 0.4625 + 0.1145

= 0.577

The confidence interval lower limit = ($\mu_{\text{Pred}}$ – CI) = 0.348

$$0.348 \leq \mu \leq 0.577$$

That is, the estimated value from the confirmation experiment is within the confidence interval. Thus, the experiment is validated.

***Source:*** Kesavan, R., Krishnaiah K., and Manikandan, N., Ergonomic Study of eye strain on video display terminal users using ANOVA Method, *Journal of Indian Institution of Industrial Engineering*, vol. XXXI, no. 11, November 2002.

## 16.4 OPTIMIZATION OF FLASH BUTT WELDING PROCESS

This study has been conducted in one of the leading boiler manufacturing company in India. One of the major problems facing the company is tube failure. The entire length of the tube consists of thousands of joints out of which the flash butt welded joints will be about 30%. Even if a single joint fails the accessibility for repair would be a serious problem and costly. So, this study has been undertaken to optimize the welding process parameters to avoid defects. As the response variable is defects, attribute data analysis is applicable.

***Selection of factors and levels:*** The factors are selected from the supplier's manual and the welding process specification sheet issued by the welding technology centre of the company. The selected parameters and their present levels are given in Table 16.26.

**TABLE 16.26** Selected parameters for Case 16.4

| *Factors* | *Present setting* |
|---|---|
| Flash length (*A*) | 6 mm |
| Upset length (*B*) | 3 mm |
| Initial die opening (*C*) | 30 mm |
| Voltage selection (*D*) | min $B_2$ |
| Welding voltage relay (*E*) | 24 |
| Forward movement (*F*) | 0.8 mm |

After discussing with the shop floor engineers it was decided to study all factors, each at three levels. The proposed settings of the factors are shown in Table 16.27.

**TABLE 16.27** Proposed levels for experimentation of Case 16.4

| *Factors* | *Level* 1 | *Level* 2 | *Level* 3 |
|---|---|---|---|
| Flash length (mm) (*A*) | 6 | 7 | 8 |
| Upset length (mm) (*B*) | 3 | 4 | 5 |
| Initial die opening (mm) (*C*) | 40 | 41 | 42 |
| Voltage selection (*D*) | min $B_3$ | $B_4$ | $B_5$ |
| Welding voltage relay (*E*) | 23 | 26 | 29 |
| Forward movement (mm) (*F*) | 0.6 | 0.9 | 1.2 |

***Conduct of experiment:*** The six factors each at three levels (12 degrees of freedom) is to be studied. Hence $L_{18}$ OA has been selected and the factors are assigned as given in Table 16.28. After conducting the experiment, the crack observed in the bend test is considered as response variable. A sample size of 90 has been selected for the whole experiment. The experiment was conducted and data are collected which is given in Table 16.28.

**TABLE 16.28** Experimental design and results for Case 16.4

| *Trial no.* | 2 *A* | 3 *B* | 4 *C* | 5 *D* | 6 *E* | 7 *F* | *Good* | *Bad* |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 1 |
| 2 | 1 | 2 | 2 | 2 | 2 | 2 | 0 | 5 |
| 3 | 1 | 3 | 3 | 3 | 3 | 3 | 1 | 4 |
| 4 | 2 | 1 | 1 | 2 | 2 | 3 | 0 | 5 |
| 5 | 2 | 2 | 2 | 3 | 3 | 1 | 1 | 4 |
| 6 | 2 | 3 | 3 | 1 | 1 | 2 | 3 | 2 |
| 7 | 3 | 1 | 2 | 1 | 3 | 2 | 3 | 2 |
| 8 | 3 | 2 | 3 | 2 | 1 | 3 | 3 | 2 |
| 9 | 3 | 3 | 1 | 3 | 2 | 1 | 4 | 1 |
| 10 | 1 | 1 | 3 | 3 | 2 | 2 | 0 | 5 |
| 11 | 1 | 2 | 1 | 1 | 3 | 3 | 2 | 3 |
| 12 | 1 | 3 | 2 | 2 | 1 | 1 | 3 | 2 |
| 13 | 2 | 1 | 2 | 3 | 1 | 3 | 5 | 0 |
| 14 | 2 | 2 | 3 | 1 | 2 | 1 | 5 | 0 |
| 15 | 2 | 3 | 1 | 2 | 3 | 2 | 3 | 2 |
| 16 | 3 | 1 | 3 | 2 | 3 | 1 | 0 | 5 |
| 17 | 3 | 2 | 1 | 3 | 1 | 2 | 0 | 5 |
| 18 | 3 | 3 | 2 | 1 | 2 | 3 | 3 | 2 |

Each trial was repeated for 5 times and in that the good one indicates those units which have passed the bend test without cracking and the bad one represents a failed unit.

***Data analysis considering the two class data as 0 and 1***

We consider good = 0 and bad = 1.

The response totals for each level of all factors is given in Table 16.29.

**TABLE 6.29** Response totals for Case 16.4

| *Factors* | *A* | *B* | *C* | *D* | *E* | *F* |
|---|---|---|---|---|---|---|
| Level 1 | 20 | 18 | 17 | 10 | 12 | 13 |
| Level 2 | 13 | 19 | 15 | 21 | 18 | 21 |
| Level 3 | 17 | 13 | 18 | 19 | 20 | 16 |

***Computation of sum of squares:***

Good = 0 and bad = 1

Grand total ($T$) = 50 (sum of all 1s. That is, bad items)

$$\text{Correction Factor (CF)} = \frac{T^2}{N} \quad \text{(N = total number of parts tested)}$$

$$= \frac{(50)^2}{90} = 27.7778$$

$$\text{Total sum of squares } (SS_{\text{Total}}) = T - \text{CF}$$

$$= 50 - 27.7778 = 22.2222$$

$$SS_A = \frac{A_1^2}{n_{A1}} + \frac{A_2^2}{n_{A2}} + \frac{A_3^2}{n_{A3}} - \text{CF}$$

$$= \frac{(20)^2}{30} + \frac{(13)^2}{30} + \frac{(17)^2}{30} - \text{CF}$$

$$= 28.6000 - 27.7778$$

$$= 0.8222$$

Similarly, other factors sum of squares are

$SS_B = 0.6888$, $SS_C = 0.1555$, $SS_D = 2.2888$, $SS_E = 1.1555$, $SS_F = 1.0888$

$SS_e = SS_{\text{Total}}$ – (sum of all factor sum of squares) = 16.0226

The Analysis of Variance is given in Tables 16.30 and 16.31.

**TABLE 16.30** ANOVA for Case 16.4

| *Source* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ | *C*(%) |
|---|---|---|---|---|---|
| *A* | 0.8222 | 2 | 0.4111 | 1.97 | 3.70 |
| *B* | 0.6888 | 2 | 0.3444 | 1.65 | 3.10 |
| *C* | 0.1555 | 2 | 0.0777 | 0.37 | 0.70 |
| *D* | 2.2888 | 2 | 1.1444 | 5.50* | 10.30 |
| *E* | 1.1555 | 2 | 0.5777 | 2.77 | 5.20 |
| *F* | 1.0888 | 2 | 0.5444 | 2.61 | 4.90 |
| Error (pure) | 16.0226 | 77 | 0.2080 | | 72.10 |
| Total | 22.2222 | 89 | | | 100.00 |

$F_{0.05,2,77} = 3.15$; *Significant

**TABLE 16.31** Pooled ANOVA for Case 16.4

| *Source* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ | *C*(%) |
|---|---|---|---|---|---|
| *D* | 2.2888 | 2 | 1.1444 | 5.37* | 10.30 |
| *E* | 1.1555 | 2 | 0.5777 | 2.71 | 5.20 |
| *F* | 1.0888 | 2 | 0.5444 | 2.55 | 4.90 |
| Pooled error | 17.6891 | 83 | 0.2131 | | 79.60 |
| Total | 22.2222 | 89 | | | 100.00 |

$F_{0.05,2,83} = 3.11$; *Significant at 5% level

***Optimal levels for the factors:*** The result shows that only one factor (*D*), namely voltage selection is significant. For the significant factor, level is selected based on minimization of average response and for the remaining factors, also the best levels are selected from the average response. The recommended levels are given in Table 16.32.

**TABLE 16.32** Recommended optimum levels for Case 16.4

| *Factors* | *Optimal level* |
|---|---|
| Flash length (*A*) | 7 mm |
| Upset length (*B*) | 5 mm |
| Initial die opening (*C*) | 40 mm |
| Voltage selection (*D*) | min B3 |
| Welding voltage relay (*E*) | 23 |
| Forward movement (*F*) | 0.6 mm |

***Predicting the optimum process average:*** For predicting the process average, one-half of the factors have been considered (Pooled ANOVA).

The optimum predicted response at the optimum condition is

$$\mu_{\text{Pred}} = \bar{Y} + (\bar{D}_1 - \bar{Y}) + (\bar{E}_1 - \bar{Y}) + (\bar{F}_1 - \bar{Y}) \tag{16.5}$$

Since the problem is concerned with attribute data, the fraction defective is computed and the omega transformation has been obtained (Table 16.33). This is used for predicting the optimum average.

**TABLE 16.33** Fraction defectives and omega transformation for Case 16.4

| *Factors* | *Fraction defective* | *Omega*($\Omega$) *value* |
|---|---|---|
| Overall mean($\bar{Y}$) | 50/90 = 0.5555 | 0.9681 |
| $\bar{D}_1$ | 10/30 = 0.3333 | –3.0109 |
| $\bar{E}_1$ | 12/30 = 0.4000 | –1.7609 |
| $\bar{F}_1$ | 13/30 = 0.4333 | –1.1656 |

$$\Omega(\text{db}) = 10 \log\left[\frac{p/100}{1 - p/100}\right] \tag{16.6}$$

The $\Omega$ value is used for predicting the optimum average.

$$\Omega_{\text{Pred}} = 0.9681 + (-3.0109 - 0.9681) + (-1.7609 - 0.9681) + (-1.1656 - 0.9681)$$
$$= -7.8736$$

Converting back omega value, $\mu_{\text{Pred}} = \dfrac{1}{1 + 10^{-\Omega/10}} = 0.1405$ (16.7)

That is, the average predicted value of $p = 0.14$.

***Confirmation experiment:*** This has been conducted with the recommended factor levels on ten units and all the ten have passed the test.

***Confidence interval for the predicted mean***

$$\text{CI} = \sqrt{F_{\alpha,v1,v2} \times MS_e \times \frac{1}{n_{\text{eff}}}} \tag{16.8}$$
$$= \sqrt{3.11 \times 0.2131 \times 0.0778}$$
$$= 0.2271$$

Therefore, the confidence interval for the optimum predicted mean is

$$\mu_{\text{Pred}} - \text{CI} \le \mu_{\text{Pred}} \le \mu_{\text{Pred}} + \text{CI}$$
$$0.1405 - 0.2271 \le \mu_{\text{Pred}} \le 0.1405 + 0.2271$$
$$-0.0866 \le \mu_{\text{Pred}} \le 0.3676$$

This indicates that the predicted average lies within the confidence interval.

***Confidence Interval for the confirmation experiment:*** From the confirmation experiment the average fraction defective is zero (no defectives observed).

$$\text{CI} = \sqrt{F_{\alpha,v1,v2} \times MS_e \times [(1/n_{\text{eff}}) + 1/r]} \tag{16.9}$$

$$= \sqrt{3.11 \times 0.2131 \times (0.0778 + 0.1)}$$

$$= 0.3433$$

The interval range is

$$-0.3423 \leq \mu_{\text{confirmation}} \leq 0.3423$$

The confirmation value lies within the confidence interval.

***Conclusion:*** The confidence interval for the predicted process average (–0.0866, 0.3676) overlaps fairly well with the confidence interval of the confirmation experiment (–0.3323, 0.3323). Thus, we can conclude that the experimental results can be reproducible.

***Source:*** Jaya, A., Improving quality of flash butt welding process using Taguchi method, Unpublished Graduation thesis Supervised by Krishnaiah K., Professor, Department of Industrial Engineering, Anna University, Chennai, India, May 2007.

## 16.5 WAVE SOLDERING PROCESS OPTIMIZATION

This case study was conducted in a company manufacturing Printed Circuit Boards (PCBs) which are used in scanners. Hence, high quality of PCBs is required. These PCBs after assembly soldered using a wave soldering machine. After soldering, each PCB is inspected for soldering defects. The present average level of defects is 6920 ppm. While zero defects are the goal, the present level of defects is very high. From the company it was learnt that the soldering machine was imported from Japan and the levels of the various parameters of the process set initially were not changed. Hence, it was decided to improve quality by optimizing the process parameters using Taguchi Method.

***Identification of factors:*** A detailed discussion was held with all the engineers concerned and the operators and decided to study the process parameters. The following process parameters (factors) were selected for study.

- Specific gravity of flux
- Preheat temperature of the board
- Solder bath temperature
- Conveyor speed

The present setting and operating range of these parameters is given in Table 16.34.

### *Selection of factors and levels*

It was decided to study all the factors at two levels. Also, decided not to disturb the routine production schedule. The process engineer suggested not to change the parameter values too far because he was doubtful about the output quality. Hence, the values are changed marginally. The proposed levels for the factors are given in Table 16.35.

**TABLE 16.34** Present setting of the factors for Case 16.5

| *Factors* | *Operating range* | *Present setting* |
|---|---|---|
| Specific gravity of flux (*A*) | 0.820–0.840 | 0.825 |
| Preheat temperature (*B*) | – | 180°C |
| Solder bath temperature (*C*) | 240°C–260°C | 246°C |
| Conveyor speed (*D*) | 0.80–1.20 m/min | 1.00 |

**TABLE 16.35** Proposed factor levels for Case 16.5

| *Factors* | *Level* 1 | *Level* 2 |
|---|---|---|
| Specific gravity of flux (*A*) | 0.822 | 0.838 |
| Preheat temperature (*B*) | 165°C | 195 |
| Solder bath temperature (*C*) | 242°C | 258°C |
| Conveyor speed (*D*) | 0.84 m/min | 0.96 |

The levels for factors *A*, *C* and *D* have been fixed as follows. For example, consider *A*. The variation in the operating range of factor $A = 0.84 - 0.82 = 0.02$

Level 1 = starting value of operating range +10% of 0.02 = 0.820 + 0.002 = 0.822

Level 2 = upper value of operating range –10% of 0.02 = 0.840 – 0.002 = 0.838

For factor *B*, the two levels are fixed as 180 ± 15, where 180 is the present setting.

In addition to the main factors, it was suspected that the two factor interactions *AB*, *AC*, *AD*, *BC*, *BD* and *CD* may have an affect on the response.

***Measurement method:*** There are about 60 varieties of PCBs differing in size and number of holes. Hence all the PCBs are classified based on the number of holes/cm$^2$ into high density and low density.

High density: more than 4 holes/cm$^2$
Low density: up to 4 holes/cm$^2$

High density PCBs are considered for this study. Whenever these types of PCBs are scheduled for production, a given experiment was conducted. Since the number of defects in a PCB is low, the following measure has been used as response variable.

$$\frac{\text{No. of defects in a PCB}}{\text{Total no. of holes}} \times 1000$$

***Design of experiment and data collection:*** To study 4 main factors and 6 two-factor interactions, $L_{16}$ OA has been selected. Hence, it will be a $2^4$ full factorial experiment. And all higher order interactions can also be analysed. The assignment of factors and interactions is given in Table 16.36. The data collected is also tabulated in Table 16.36. Note that the response (*Y*) given is the sum of two replications.

**TABLE 16.36** Assignment of factors and interactions for Case 16.5 ($L_{16}$ OA)

| *Trial no.* | *Columns/Factors and interactions* | | | | | | | | | | | | | | | *Y* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | |
| | *C* | *B* | *BC* | *D* | *CD* | *BD* | *BCD* | *ABCD* | *ABD* | *ACD* | *AD* | *ABC* | *AB* | *AC* | *A* | |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 13.5 |
| 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 7.5 |
| 3 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 16.9 |
| 4 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 9.6 |
| 5 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 3.2 |
| 6 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 29.7 |
| 7 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 11.7 |
| 8 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 7.4 |
| 9 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 6.9 |
| 10 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 6.9 |
| 11 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 14.6 |
| 12 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 12.3 |
| 13 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 12.0 |
| 14 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 14.8 |
| 15 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 11.0 |
| 16 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 10.0 |

$Y$ = Response = Sum of two replications

***Data analysis:*** The data is analysed considering the response as variable data. The response totals for all the factors and interactions are given in Table 16.37.

***Computation of sum of squares:***

Grand total ($T$) = 188

Total number of observations ($N$) = 32

$$\text{Overall mean } (\bar{Y}) = \frac{188}{32} = 5.875$$

$$\text{Correction factor (CF)} = \frac{T^2}{N} = \frac{(188)^2}{32}$$

$$SS_{\text{Total}} = (5.4)^2 + (8.1)^2 + \ldots + (5.7)^2 - \frac{(188)^2}{32}$$

$$= 303.88$$

$$SS_A = \frac{A_1^2}{n_{A1}} + \frac{A_2^2}{n_{A2}} - \text{CF}$$

$$= \frac{(108)^2}{16} + \frac{(80)^2}{16} - \frac{(188)^2}{32}$$

$$= 24.50$$

Similarly, the sum of squares of all effects are computed and summarized in the initial ANOVA (Table 16.38). After pooling the sum of squares as usual, the final ANOVA is given in Table 16.39.

**TABLE 16.37** Response totals for Case 16.5

| *Factor effects* | *Level* 1 | *Level* 2 | *Factor effects* | *Level* 1 | *Level* 2 |
|---|---|---|---|---|---|
| *A* | 108 | 80.0 | *BD* | 74.9 | 113.1 |
| *B* | 88.2 | 99.8 | *CD* | 101.8 | 86.2 |
| *C* | 99.5 | 88.5 | *ABC* | 74.9 | 113.1 |
| *D* | 94.5 | 93.5 | *ABD* | 89.3 | 98.7 |
| *AB* | 81 | 107.0 | *ACD* | 113.8 | 74.2 |
| *AC* | 109.5 | 78.5 | *BCD* | 93.8 | 94.2 |
| *AD* | 109.7 | 78.3 | *ABCD* | 89.8 | 98.2 |
| *BC* | 95.3 | 92.7 | | | |

**TABLE 16.38** Initial ANOVA for Case 16.5

| *Source* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ |
|---|---|---|---|---|
| *A* | 24.500 | 1 | 24.500 | 10.77 |
| *B* | 4.205 | 1 | 4.205 | 1.85 |
| *C* | 3.781 | 1 | 3.781 | 9.28 |
| *D* | 0.031 | 1 | 0.031 | 1.66 |
| *AB* | 21.125 | 1 | 21.125 | 13.20 |
| *AC* | 30.031 | 1 | 30.031 | 0.09 |
| *AD* | 30.811 | 1 | 30.811 | 0.01 |
| *BC* | 0.211 | 1 | 0.211 | 13.54 |
| *BD* | 45.601 | 1 | 45.601 | 20.04 |
| *CD* | 7.605 | 1 | 7.605 | 3.34 |
| *ABC* | 45.601 | 1 | 45.601 | 0.002 |
| *ABD* | 2.761 | 1 | 2.761 | 1.21 |
| *ACD* | 49.005 | 1 | 49.005 | 21.54 |
| *BCD* | 0.005 | 1 | 0.005 | 20.04 |
| *ABCD* | 2.205 | 1 | 2.205 | 0.97 |
| Error | 36.402 | 16 | 2.275 | |
| Total | 303.880 | 31 | | |

**TABLE 16.39** Final ANOVA for Case 16.5

| *Source* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ | *C*(%) |
|---|---|---|---|---|---|
| *A* | 24.500 | 1 | 24.500 | 11.34* | 8.06 |
| *AB* | 21.125 | 1 | 21.125 | 9.78* | 6.95 |
| *AC* | 30.031 | 1 | 30.031 | 13.90* | 9.88 |
| *AD* | 30.811 | 1 | 30.811 | 14.26* | 10.14 |
| *BD* | 45.601 | 1 | 45.601 | 21.11* | 15.01 |
| *CD* | 7.605 | 1 | 7.605 | 3.52 | 2.5 |
| *ABC* | 45.601 | 1 | 45.601 | 22.69* | 15.01 |
| *ACD* | 49.005 | 1 | 49.005 | 21.11* | 16.13 |
| Pooled error | 49.601 | 23 | 2.16 | | 16.32 |
| Total | 303.88 | 31 | | | 100.00 |

$F_{0.05,1,23}$ = 4.28; *Significant

***Optimum levels for factors:*** From Table 16.39, it is found that only one main factor (*A*), the two-factor interactions (*AB*, *AC*, *AD*, and *BD*) and the three-factor interactions *ABC* and *ACD* are significant. Hence for determining the optimum levels for the factors the average response for all possible combinations of these effects are evaluated and given in Table 16.40. Since the objective is to minimize the response (measure of defects), the combinations $\mathbf{A_2C_1D_1 = 2.675}$ and $\mathbf{A_2B_2C_1 = 2.675}$ are selected (Table 16.40). We have to find the optimal levels for all the four factors *A*, *B*, *C* and *D*. Since *B* is not present in $\mathbf{A_2C_1D_1}$ and *D* is not in $A_2B_2C_1$, the mean values of the following combinations are obtained.

$$A_2C_1D_1B_1 = 3.75$$
$$A_2C_1D_1B_2 = 1.60$$
$$A_2B_2C_1D_1 = 1.60$$
$$A_2B_2C_1D_2 = 3.70$$

Therefore, the optimal combination is $\mathbf{A_2C_1D_1B_2}$. The existing and proposed (optimum) levels are given in Table 16.41.

**TABLE 16.40** Mean values of significant effects for Case 16.5

| $A_1$ = 6.75 | | $A_2$ = 5.00 | |
|---|---|---|---|
| $A_1B_1$ = 5.575 | $A_1C_1$ = 8.063 | $A_1D_1$ = 7.763 | $B_1D_1$ = 4.350 |
| $A_1B_2$ = 7.925 | $A_1C_2$ = 5.438 | $A_1D_2$ = 5.738 | $B_1D_2$ = 6.675 |
| $A_2B_1$ = 5.450 | $A_2C_1$ = 4.375 | $A_2D_1$ = 4.050 | $B_2D_1$ = 7.463 |
| $A_2B_2$ = 4.550 | $A_2C_2$ = 5.625 | $A_2D_2$ = 5.950 | $B_2D_2$ = 5.012 |
| $A_1C_1D_1$ = 10.800 | $\mathbf{A_2C_1D_1 = 2.675}$ | $A_1B_1C_1$ = 5.775 | $A_2B_1C_1$ = 6.100 |
| $A_1C_1D_2$ = 5.325 | $A_2C_1D_2$ = 6.075 | $A_1B_1C_2$ = 5.375 | $A_2B_1C_2$ = 4.800 |
| $A_1C_2D_1$ = 4.725 | $A_2C_2D_1$ = 5.425 | $A_1B_2C_1$ = 10.350 | $\mathbf{A_2B_2C_1 = 2.675}$ |
| $A_1C_2D_2$ = 6.150 | $A_2C_2D_2$ = 5.825 | $A_1B_2C_2$ = 5.500 | $A_2B_2C_2$ = 6.450 |

The existing and proposed (optimum) levels are given below (Table 16.41)

**TABLE 16.41** Existing and Proposed settings for Case 16.5

| *Factors* | *Existing Level* | *Proposed Level* |
|---|---|---|
| Specific gravity of flux ($A$) | 0.825 | 0.838 |
| Preheat temperature ($B$) | 180°C | 195°C |
| Solder bath temperature ($C$) | 246°C | 242°C |
| Conveyor speed ($D$) | 1.0 m/min | 0.84 m/min |

$$\text{Predicted optimum response } (\mu_{\text{Pred}}) = \bar{Y} + (A_2B_2C_1D_1 - \bar{Y}) \quad (16.10)$$
$$= 5.875 + (1.60 - 5.875)$$
$$= 1.60$$

***Confirmation experiment:*** A confirmation experiment with two replications at the optimal factor settings was conducted. It was found that the defects reduced from 6920 ppm to 2365, a reduction of 65%.

***Source:*** Krishnaiah, K., Taguchi Approach to Process Improvement—A Case Study, National Conference on Industrial Engineering Towards 21st Century, Organized by the Dept. of Mechanical Engineering, S.V. University, Tirupati, Jan. 31–Feb. 2, 1998.

## 16.6 APPLICATION OF $L_{27}$ OA

A study was conducted on induction hardening heat treatment process. The three induction hardening parameters, namely, power potential ($A$), scan speed ($B$) and quench flow rate ($C$) have been selected and decided to study all factors at three levels. In addition, it was suspected that the two factor interactions $AB$, $AC$ and $BC$ may also influence the surface hardness and hence these are also included in the study. Table 16.42 shows the control parameters and their levels.

**TABLE 16.42** Factors and levels for Case 16.6

| *Factors* | *Level* 1 | *Level* 2 | *Level* 3 |
|---|---|---|---|
| Power potential (kW/inch$^2$) ($A$) | 6.5 | 7.5 | 8.5 |
| Scan speed (m/min) ($B$) | 1.5 | 2.0 | 2.5 |
| Quench flow rate (l/min) ($C$) | 15.0 | 17.5 | 20.0 |

***Design of experiment:*** Since the study involves three main factors each at three levels and three two factor interactions, $L_{27}$ OA has been selected. The following linear graph (Figure 16.5) was used to assign the main factors and interactions.

The experimental design layout is given in Table 16.43. Since each interaction has four degrees of freedom, the interaction is assigned to two columns. The vacant columns are assigned by $e$, indicating error.

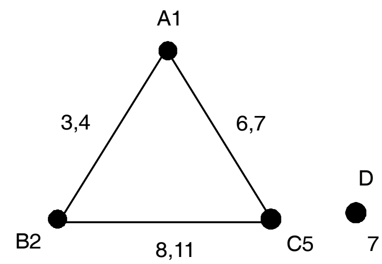



**FIGURE 16.5** Required linear graph for the Case 16.6.

***Data collection:*** The order of experimentation was random. Only one replication was made. The surface hardness ($Y$) measured is given in the last column of Table 16.43.

**TABLE 16.43** Experimental design layout and data for Case 16.6 ($L_{27}$ OA)

| *Trial no.* | 1 *A* | 2 *B* | 3 *AB* | 4 *AB* | 5 *C* | 6 *AC* | 7 *AC* | 8 *BC* | 9 *e* | 10 *e* | 11 *BC* | 12 *e* | 13 *e* | *Y* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 80 |
| 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 80 |
| 3 | 1 | 1 | 1 | 1 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 77 |
| 4 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 77 |
| 5 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 1 | 1 | 1 | 82 |
| 6 | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 1 | 1 | 1 | 2 | 2 | 2 | 81 |
| 7 | 1 | 3 | 3 | 3 | 1 | 1 | 1 | 3 | 3 | 3 | 2 | 2 | 2 | 80 |
| 8 | 1 | 3 | 3 | 3 | 2 | 2 | 2 | 1 | 1 | 1 | 3 | 3 | 3 | 79 |
| 9 | 1 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 1 | 1 | 1 | 79 |
| 10 | 2 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 80 |
| 11 | 2 | 1 | 2 | 3 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 79 |
| 12 | 2 | 1 | 2 | 3 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 75 |
| 13 | 2 | 2 | 3 | 1 | 1 | 2 | 3 | 2 | 3 | 1 | 3 | 1 | 2 | 80 |
| 14 | 2 | 2 | 3 | 1 | 2 | 3 | 1 | 3 | 1 | 2 | 1 | 2 | 3 | 78 |
| 15 | 2 | 2 | 3 | 1 | 3 | 1 | 2 | 1 | 2 | 3 | 2 | 3 | 1 | 80 |
| 16 | 2 | 3 | 1 | 2 | 1 | 2 | 3 | 3 | 1 | 2 | 2 | 3 | 1 | 81 |
| 17 | 2 | 3 | 1 | 2 | 2 | 3 | 1 | 1 | 2 | 3 | 3 | 1 | 2 | 74 |
| 18 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 2 | 3 | 1 | 1 | 2 | 3 | 79 |
| 19 | 3 | 1 | 3 | 2 | 1 | 3 | 2 | 1 | 3 | 2 | 1 | 3 | 2 | 66 |
| 20 | 3 | 1 | 3 | 2 | 2 | 1 | 3 | 2 | 1 | 3 | 2 | 1 | 3 | 62 |
| 21 | 3 | 1 | 3 | 2 | 3 | 2 | 1 | 3 | 2 | 1 | 3 | 2 | 1 | 59 |
| 22 | 3 | 2 | 1 | 3 | 1 | 3 | 2 | 2 | 1 | 3 | 3 | 2 | 1 | 68 |
| 23 | 3 | 2 | 1 | 3 | 2 | 1 | 3 | 3 | 2 | 1 | 1 | 3 | 2 | 65 |
| 24 | 3 | 2 | 1 | 3 | 3 | 2 | 1 | 1 | 3 | 2 | 2 | 1 | 3 | 65 |
| 25 | 3 | 3 | 2 | 1 | 1 | 3 | 2 | 3 | 2 | 1 | 2 | 1 | 3 | 61 |
| 26 | 3 | 3 | 2 | 1 | 2 | 1 | 3 | 1 | 3 | 2 | 3 | 2 | 1 | 64 |
| 27 | 3 | 3 | 2 | 1 | 3 | 2 | 1 | 2 | 1 | 3 | 1 | 3 | 2 | 61 |

***Data analysis:*** The response totals for all the factors, interactions and the vacant columns are given in Table 16.44.

**TABLE 16.44** Response (hardness) totals for all effects of Case 16.6

| *Factor effect/Vacant col.* | *Level* 1 | *Level* 2 | *Level* 3 |
|---|---|---|---|
| $A$ | 715 | 706 | 571 |
| $B$ | 658 | 676 | 658 |
| $C$ | 673 | 663 | 656 |
| $AB$: Col. 3 | 669 | 660 | 663 |
| Col. 4 | 661 | 661 | 670 |
| $AC$: Col. 6 | 662 | 667 | 663 |
| Col. 7 | 653 | 670 | 669 |
| $BC$: Col. 8 | 669 | 665 | 658 |
| Col. 11 | 670 | 669 | 653 |
| $e_9$ | 665 | 655 | 672 |
| $e_{10}$ | 663 | 665 | 664 |
| $e_{12}$ | 658 | 669 | 665 |
| $e_{13}$ | 672 | 662 | 658 |

***Computation of sum of squares:***

Grand total = 1992

Total number of observation = 27

$$\text{Correction factor (CF)} = \frac{(1992)^2}{27} = 146965.33$$

$$SS_{\text{Total}} = (80)^2 + (80)^2 + (77)^2 + \cdots + (61)^2 - \text{CF}$$

$$= 1580.67$$

$$SS_A = \frac{A_1^2}{n_{A1}} + \frac{A_2^2}{n_{A2}} + \frac{A_3^2}{n_{A3}} - \text{CF}$$

$$= \frac{(715)^2}{9} + \frac{(706)^2}{9} + \frac{(571)^2}{9} - \text{CF}$$

$$= 1446.00$$

Similarly, the other sum of squares is calculated (main effects and vacant columns). The sum of squares of interaction, say $AB$ is the sum of $SS_{\text{col. 3}}$ and $SS_{\text{col. 4}}$.

$SS_B = 24.00$, $SS_C = 16.22$, $SS_{AB} = SS_{\text{col. 3}} + SS_{\text{col. 4}} = 4.66 + 6.00 = 10.66$, $SS_{AC} = SS_{\text{col. 6}} + SS_{\text{col. 7}} = 1.55 + 20.22 = 21.77$, $SS_{BC} = SS_{\text{col. 8}} + SS_{\text{col. 11}} = 6.88 + 20.22 = 27.10$, $SS_{e_9} = 16.23$, $SS_{e_{10}} = 0.23$, $SS_{e_{12}} = 6.88$, $SS_{e_{13}} = 11.58$.

These sum of squares are given in the initial ANOVA Table 16.45.

From Table 16.45, it is seen that the interactions are not significant and their contribution to the variation is also negligible. These interactions are pooled into the error term and the final ANOVA is given in Table 16.46. Even after pooling it is seen that the main effect $A$ alone is significant.

**TABLE 16.45** ANOVA (initial) for Case 16.6

| *Source* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ | $C(\%)$ |
|---|---|---|---|---|---|
| $A$ | 1446.00 | 2 | 723.00 | 165.44* | 91.48 |
| $B$ | 24.00 | 2 | 12.00 | 2.74 | 1.52 |
| $C$ | 16.22 | 2 | 8.11 | 1.85 | 1.03 |
| $AB$ | 10.66 | 4 | 2.67 | 0.61 | 0.67 |
| $AC$ | 21.77 | 4 | 5.44 | 1.24 | 1.37 |
| $BC$ | 27.10 | 4 | 6.78 | 1.55 | 1.72 |
| Error (vacant columns) | 34.92 | 8 | 4.37 | | 2.21 |
| Total | 1580.67 | 26 | | | 100.00 |

$F_{0.05,2,8} = 4.46$ $F_{0.05,4,8} = 3.84$; *Significant

Since the objective is to maximize the hardness, the optimal levels for the factors based on maximum average response are $A_1$, $B_2$ and $C_1$ (Table 16.44).

**TABLE 16.46** Final ANOVA with pooled error for Case 16.6

| *Source* | *Sum of squares* | *Degrees of freedom* | *Mean square* | $F_0$ | $C(\%)$ |
|---|---|---|---|---|---|
| $A$ | 1446.00 | 2 | 723.00 | 153.17 | 91.48 |
| $B$ | 24.00 | 2 | 12.00 | 2.54 | 1.52 |
| $C$ | 16.22 | 2 | 8.11 | 1.71 | 1.03 |
| Pooled error | 94.45 | 20 | 4.72 | | 5.97 |
| Total | 1580.67 | 26 | | | 100.00 |

$F_{0.05,2,20} = 3.49$; *Significant

***Conclusion:*** This case shows how interactions between three level factors can be handled in Taguchi experiments.

APPENDIX

# A

## A.1 STANDARD NORMAL DISTRIBUTION TABLE

| *z* | *0.00* | *0.01* | *0.02* | *0.03* | *0.04* | *0.05* | *0.06* | *0.07* | *0.08* | *0.09* |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0000 | 0040 | 0080 | 0120 | 0160 | 0199 | 0239 | 0279 | 0319 | 0359 |
| 0.1 | 0398 | 0438 | 0478 | 0517 | 0557 | 0596 | 0636 | 0675 | 0714 | 0753 |
| 0.2 | 0793 | 0832 | 0871 | 0910 | 0948 | 0987 | 1026 | 1064 | 1103 | 1141 |
| 0.3 | 1179 | 1217 | 1255 | 1293 | 1331 | 1368 | 1406 | 1443 | 1480 | 1517 |
| 0.4 | 1554 | 1591 | 1628 | 1664 | 1700 | 1736 | 1772 | 1808 | 1844 | 1879 |
| 0.5 | 1915 | 1950 | 1985 | 2019 | 2054 | 2088 | 2123 | 2157 | 2190 | 2224 |
| 0.6 | 2257 | 2291 | 2324 | 2357 | 2389 | 2422 | 2454 | 2486 | 2517 | 2549 |
| 0.7 | 2580 | 2611 | 2642 | 2673 | 2704 | 2734 | 2764 | 2794 | 2823 | 2852 |
| 0.8 | 2881 | 2910 | 2939 | 2967 | 2995 | 3023 | 3051 | 3078 | 3106 | 3133 |
| 0.9 | 3159 | 3186 | 3212 | 3238 | 3264 | 3289 | 3315 | 3340 | 3365 | 3389 |
| 1.0 | 3413 | 3438 | 3461 | 3485 | 3508 | 3531 | 3554 | 3577 | 3599 | 3621 |
| 1.1 | 3643 | 3665 | 3686 | 3708 | 3729 | 3749 | 3770 | 3790 | 3810 | 3830 |
| 1.2 | 3849 | 3869 | 3888 | 3907 | 3925 | 3944 | 3962 | 3980 | 3997 | 4015 |
| 1.3 | 4032 | 4049 | 4066 | 4082 | 4099 | 4115 | 4131 | 4147 | 4162 | 4177 |
| 1.4 | 4192 | 4207 | 4222 | 4236 | 4251 | 4265 | 4279 | 4292 | 4306 | 4319 |
| 1.5 | 4332 | 4345 | 4357 | 4370 | 4382 | 4394 | 4406 | 4418 | 4429 | 4441 |
| 1.6 | 4452 | 4463 | 4474 | 4484 | 4495 | 4505 | 4515 | 4525 | 4535 | 4545 |
| 1.7 | 4554 | 4564 | 4573 | 4582 | 4591 | 4599 | 4608 | 4616 | 4625 | 4633 |
| 1.8 | 4641 | 4649 | 4656 | 4664 | 4671 | 4678 | 4686 | 4693 | 4699 | 4706 |
| 1.9 | 4713 | 4719 | 4726 | 4732 | 4738 | 4744 | 4750 | 4756 | 4761 | 4767 |
| 2.0 | 4772 | 4778 | 4783 | 4788 | 4793 | 4798 | 4803 | 4808 | 4812 | 4817 |
| 2.1 | 4821 | 4826 | 4830 | 4834 | 4838 | 4842 | 4846 | 4850 | 4854 | 4857 |
| 2.2 | 4861 | 4864 | 4868 | 4871 | 4875 | 4878 | 4881 | 4884 | 4887 | 4890 |
| 2.3 | 4893 | 4896 | 4898 | 4901 | 4904 | 4906 | 4909 | 4911 | 4913 | 4916 |
| 2.4 | 4918 | 4920 | 4922 | 4925 | 4927 | 4929 | 4931 | 4932 | 4934 | 4936 |
| 2.5 | 4938 | 4940 | 4941 | 4943 | 4945 | 4946 | 4948 | 4949 | 4951 | 4952 |

(*Contd.*)

| z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 2.6 | 4953 | 4955 | 4956 | 4957 | 4959 | 4960 | 4961 | 4962 | 4963 | 4964 |
| 2.7 | 4965 | 4966 | 4967 | 4968 | 4969 | 4970 | 4971 | 4972 | 4973 | 4974 |
| 2.8 | 4974 | 4975 | 4976 | 4977 | 4977 | 4978 | 4979 | 4979 | 4980 | 4981 |
| 2.9 | 4981 | 4982 | 4982 | 4983 | 4984 | 4984 | 4985 | 4985 | 4986 | 4986 |
| 3.0 | 4987 | 4987 | 4987 | 4988 | 4988 | 4989 | 4989 | 4989 | 4990 | 4990 |

## A.2 THE *t* DISTRIBUTION TABLE

| | *Area in the right tail under the t distribution curve* | | | | | |
|---|---|---|---|---|---|---|
| *df* | *0.10* | *0.05* | *0.025* | *0.01* | *0.005* | *0.001* |
| 1 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | 318.309 |
| 2 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 22.327 |
| 3 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 10.215 |
| 4 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 7.173 |
| 5 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5.893 |
| 6 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 5.208 |
| 7 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 4.785 |
| 8 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 4.501 |
| 9 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 4.297 |
| 10 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.144 |
| 11 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 4.025 |
| 12 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 3.930 |
| 13 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 3.852 |
| 14 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 3.787 |
| 15 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 3.733 |
| 16 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 3.686 |
| 17 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 3.646 |
| 18 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 3.610 |
| 19 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.579 |
| 20 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.552 |
| 21 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | 3.527 |
| 22 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 3.505 |
| 23 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | 3.485 |
| 24 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 3.467 |
| 25 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | 3.450 |
| 26 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 3.435 |
| 27 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 3.421 |
| 28 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 3.408 |
| 29 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | 3.396 |
| 30 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 | 3.385 |
| 40 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 | 3.307 |
| 60 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 | 3.232 |
| 120 | 1.289 | 1.658 | 1.980 | 2.358 | 2.617 | 3.160 |
| $\infty$ | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 |

## A.3 PERCENTAGE POINTS OF THE $F$ DISTRIBUTION: $F_{0.05,\nu_1,\nu_2}$ ($\nu_1$ = Numerator $df$ and $\nu_2$ = denominator $df$)

| $\nu_1 \backslash \nu_2$ | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 161.5 | 199.5 | 215.7 | 224.6 | 230.2 | 234.0 | 236.8 | 238.9 | 240.5 | 241.9 |
| 2 | 18.51 | 19.00 | 19.16 | 19.25 | 19.30 | 19.33 | 19.35 | 19.37 | 19.38 | 19.40 |
| 3 | 10.13 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.89 | 8.85 | 8.81 | 8.79 |
| 4 | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.09 | 6.04 | 6.00 | 5.96 |
| 5 | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.88 | 4.82 | 4.77 | 4.74 |
| 6 | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.21 | 4.15 | 4.10 | 4.06 |
| 7 | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.79 | 3.73 | 3.68 | 3.64 |
| 8 | 5.32 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.50 | 3.44 | 3.39 | 3.35 |
| 9 | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.29 | 3.23 | 3.18 | 3.14 |
| 10 | 4.96 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.14 | 3.07 | 3.02 | 2.98 |
| 11 | 4.84 | 3.98 | 3.59 | 3.36 | 3.20 | 3.09 | 3.01 | 2.95 | 2.90 | 2.85 |
| 12 | 4.75 | 3.89 | 3.49 | 3.26 | 3.11 | 3.00 | 2.91 | 2.85 | 2.80 | 2.75 |
| 13 | 4.67 | 3.81 | 3.41 | 3.18 | 3.03 | 2.92 | 2.83 | 2.77 | 2.71 | 2.67 |
| 14 | 4.60 | 3.74 | 3.34 | 3.11 | 2.96 | 2.85 | 2.76 | 2.70 | 2.65 | 2.60 |
| 15 | 4.54 | 3.68 | 3.29 | 3.06 | 2.90 | 2.79 | 2.71 | 2.64 | 2.59 | 2.54 |
| 16 | 4.49 | 3.63 | 3.24 | 3.01 | 2.85 | 2.74 | 2.66 | 2.59 | 2.54 | 2.49 |
| 17 | 4.45 | 3.59 | 3.20 | 2.96 | 2.81 | 2.70 | 2.61 | 2.55 | 2.49 | 2.45 |
| 18 | 4.41 | 3.55 | 3.16 | 2.93 | 2.77 | 2.66 | 2.58 | 2.51 | 2.46 | 2.41 |
| 19 | 4.38 | 3.52 | 3.13 | 2.90 | 2.74 | 2.63 | 2.54 | 2.48 | 2.42 | 2.38 |
| 20 | 4.35 | 3.49 | 3.10 | 2.87 | 2.71 | 2.60 | 2.51 | 2.45 | 2.39 | 2.35 |
| 21 | 4.32 | 3.47 | 3.07 | 2.84 | 2.68 | 2.57 | 2.49 | 2.42 | 2.37 | 2.32 |
| 22 | 4.30 | 3.44 | 3.05 | 2.82 | 2.66 | 2.55 | 2.46 | 2.40 | 2.34 | 2.30 |
| 23 | 4.28 | 3.42 | 3.03 | 2.80 | 2.64 | 2.53 | 2.44 | 2.37 | 2.32 | 2.27 |
| 24 | 4.26 | 3.40 | 3.01 | 2.78 | 2.62 | 2.51 | 2.42 | 2.36 | 2.30 | 2.25 |
| 25 | 4.24 | 3.39 | 2.99 | 2.76 | 2.60 | 2.49 | 2.40 | 2.34 | 2.28 | 2.24 |
| 30 | 4.17 | 3.32 | 2.92 | 2.69 | 2.53 | 2.42 | 2.33 | 2.27 | 2.21 | 2.16 |
| 40 | 4.08 | 3.23 | 2.84 | 2.61 | 2.45 | 2.34 | 2.25 | 2.18 | 2.12 | 2.08 |
| 50 | 4.03 | 3.18 | 2.79 | 2.56 | 2.40 | 2.29 | 2.20 | 2.13 | 2.07 | 2.03 |
| 100 | 3.94 | 3.09 | 2.70 | 2.46 | 2.31 | 2.19 | 2.10 | 2.03 | 1.97 | 1.93 |

(*Contd.*)

| $v_1 \backslash v_2$ | *11* | *12* | *15* | *20* | *25* | *30* | *40* | *50* | *100* |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 243.0 | 243.9 | 246.0 | 248.0 | 249.3 | 250.1 | 251.1 | 251.8 | 253.0 |
| 2 | 19.40 | 19.41 | 19.43 | 19.45 | 19.46 | 19.46 | 19.47 | 19.48 | 19.49 |
| 3 | 8.76 | 8.74 | 8.70 | 8.66 | 8.63 | 8.62 | 8.59 | 8.58 | 8.55 |
| 4 | 5.94 | 5.91 | 5.86 | 5.80 | 5.77 | 5.75 | 5.72 | 5.70 | 5.66 |
| 5 | 4.70 | 4.68 | 4.62 | 4.56 | 4.52 | 4.50 | 4.46 | 4.44 | 4.41 |
| 6 | 4.03 | 4.00 | 3.94 | 3.87 | 3.83 | 3.81 | 3.77 | 3.75 | 3.71 |
| 7 | 3.60 | 3.57 | 3.51 | 3.44 | 3.40 | 3.38 | 3.34 | 3.32 | 3.27 |
| 8 | 3.31 | 3.28 | 3.22 | 3.15 | 3.11 | 3.08 | 3.04 | 3.02 | 2.97 |
| 9 | 3.10 | 3.07 | 3.01 | 2.94 | 2.89 | 2.86 | 2.83 | 2.80 | 2.76 |
| 10 | 2.94 | 2.91 | 2.85 | 2.77 | 2.73 | 2.70 | 2.66 | 2.64 | 2.59 |
| 11 | 2.82 | 2.79 | 2.72 | 2.65 | 2.60 | 2.57 | 2.53 | 2.51 | 2.46 |
| 12 | 2.72 | 2.69 | 2.62 | 2.54 | 2.50 | 2.47 | 2.43 | 2.40 | 2.35 |
| 13 | 2.63 | 2.60 | 2.53 | 2.46 | 2.41 | 2.38 | 2.34 | 2.31 | 2.26 |
| 14 | 2.57 | 2.53 | 2.46 | 2.39 | 2.34 | 2.31 | 2.27 | 2.24 | 2.19 |
| 15 | 2.51 | 2.48 | 2.40 | 2.33 | 2.28 | 2.25 | 2.20 | 2.18 | 2.12 |
| 16 | 2.46 | 2.42 | 2.35 | 2.28 | 2.23 | 2.19 | 2.15 | 2.12 | 2.07 |
| 17 | 2.41 | 2.38 | 2.31 | 2.23 | 2.18 | 2.15 | 2.10 | 2.08 | 2.02 |
| 18 | 2.37 | 2.34 | 2.27 | 2.19 | 2.14 | 2.11 | 2.06 | 2.04 | 1.98 |
| 19 | 2.34 | 2.31 | 2.23 | 2.16 | 2.11 | 2.07 | 2.03 | 2.00 | 1.94 |
| 20 | 2.31 | 2.28 | 2.20 | 2.12 | 2.07 | 2.04 | 1.99 | 1.97 | 1.91 |
| 21 | 2.28 | 2.25 | 2.18 | 2.10 | 2.05 | 2.01 | 1.96 | 1.94 | 1.88 |
| 22 | 2.26 | 2.23 | 2.15 | 2.07 | 2.02 | 1.97 | 1.94 | 1.91 | 1.85 |
| 23 | 2.24 | 2.20 | 2.13 | 2.05 | 2.00 | 1.96 | 1.91 | 1.88 | 1.82 |
| 24 | 2.22 | 2.18 | 2.16 | 2.03 | 1.97 | 1.94 | 1.89 | 1.86 | 1.80 |
| 25 | 2.20 | 2.16 | 2.09 | 2.01 | 1.96 | 1.92 | 1.87 | 1.84 | 1.78 |
| 30 | 2.13 | 2.09 | 2.01 | 1.93 | 1.88 | 1.84 | 1.79 | 1.76 | 1.70 |
| 40 | 2.04 | 2.00 | 1.92 | 1.84 | 1.78 | 1.74 | 1.69 | 1.66 | 1.59 |
| 50 | 1.99 | 1.95 | 1.87 | 1.78 | 1.73 | 1.69 | 1.63 | 1.60 | 1.52 |
| 100 | 1.89 | 1.85 | 1.77 | 1.68 | 1.62 | 1.57 | 1.52 | 1.48 | 1.39 |

## A.3 PERCENTAGE POINTS OF THE *F* DISTRIBUTION: $F_{0.025,\nu_1,\nu_2}$ ($\nu_1$ = Numerator *df* and $\nu_2$ = denominator *df*)

| $\nu_1 \backslash \nu_2$ | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 647.8 | 799.5 | 864.2 | 899.6 | 921.8 | 937.1 | 948.2 | 956.7 | 963.3 | 968.6 |
| 2 | 38.51 | 39.00 | 39.17 | 39.25 | 39.30 | 39.33 | 39.36 | 39.37 | 39.39 | 39.40 |
| 3 | 17.44 | 16.04 | 15.44 | 15.10 | 14.88 | 14.73 | 14.62 | 14.54 | 14.47 | 14.42 |
| 4 | 12.22 | 10.65 | 9.98 | 9.60 | 9.36 | 9.20 | 9.07 | 8.98 | 8.90 | 8.84 |
| 5 | 10.01 | 8.43 | 7.76 | 7.39 | 7.15 | 6.98 | 6.85 | 6.76 | 6.68 | 6.62 |
| 6 | 8.81 | 7.26 | 6.60 | 6.23 | 5.99 | 5.82 | 5.70 | 5.60 | 5.52 | 5.46 |
| 7 | 8.07 | 6.54 | 5.89 | 5.52 | 5.29 | 5.12 | 4.99 | 4.90 | 4.82 | 4.76 |
| 8 | 7.57 | 6.06 | 5.42 | 5.05 | 4.82 | 4.65 | 4.53 | 4.43 | 4.36 | 4.30 |
| 9 | 7.21 | 5.71 | 5.08 | 4.72 | 4.48 | 4.32 | 4.20 | 4.10 | 4.03 | 3.96 |
| 10 | 6.94 | 5.46 | 4.83 | 4.47 | 4.24 | 4.07 | 3.95 | 3.85 | 3.78 | 3.72 |
| 11 | 6.72 | 5.26 | 4.63 | 4.28 | 4.04 | 3.88 | 3.76 | 3.66 | 3.59 | 3.53 |
| 12 | 6.55 | 5.10 | 4.47 | 4.12 | 3.89 | 3.73 | 3.61 | 3.51 | 3.44 | 3.37 |
| 13 | 6.41 | 4.97 | 4.35 | 4.00 | 3.77 | 3.60 | 3.48 | 3.39 | 3.31 | 3.25 |
| 14 | 6.30 | 4.86 | 4.24 | 3.89 | 3.66 | 3.50 | 3.38 | 3.29 | 3.21 | 3.15 |
| 15 | 6.20 | 4.77 | 4.15 | 3.80 | 3.58 | 3.41 | 3.29 | 3.20 | 3.12 | 3.06 |
| 16 | 6.12 | 4.69 | 4.08 | 3.73 | 3.50 | 3.34 | 3.22 | 3.12 | 3.05 | 2.99 |
| 17 | 6.04 | 4.62 | 4.01 | 3.66 | 3.44 | 3.28 | 3.16 | 3.06 | 2.98 | 2.92 |
| 18 | 5.98 | 4.56 | 3.95 | 3.61 | 3.38 | 3.22 | 3.10 | 3.01 | 2.93 | 2.87 |
| 19 | 5.92 | 4.51 | 3.90 | 3.56 | 3.33 | 3.17 | 3.05 | 2.96 | 2.88 | 2.82 |
| 20 | 5.87 | 4.46 | 3.86 | 3.51 | 3.29 | 3.13 | 3.01 | 2.91 | 2.84 | 2.77 |
| 21 | 5.83 | 4.42 | 3.82 | 3.48 | 3.25 | 3.09 | 2.97 | 2.87 | 2.80 | 2.73 |
| 22 | 5.79 | 4.38 | 3.78 | 3.44 | 3.22 | 3.05 | 2.93 | 2.84 | 2.76 | 2.70 |
| 23 | 5.75 | 4.35 | 3.75 | 3.41 | 3.18 | 3.02 | 2.90 | 2.81 | 2.73 | 2.67 |
| 24 | 5.72 | 4.32 | 3.72 | 3.38 | 3.15 | 2.99 | 2.87 | 2.78 | 2.70 | 2.64 |
| 25 | 5.69 | 4.29 | 3.69 | 3.35 | 3.13 | 2.97 | 2.85 | 2.75 | 2.68 | 2.61 |
| 30 | 5.57 | 4.18 | 3.59 | 3.25 | 3.03 | 2.87 | 2.75 | 2.65 | 2.57 | 2.51 |
| 40 | 5.42 | 4.05 | 3.46 | 3.13 | 2.90 | 2.74 | 2.62 | 2.53 | 2.45 | 2.39 |
| 50 | 5.34 | 3.97 | 3.39 | 3.05 | 2.83 | 2.67 | 2.55 | 2.46 | 2.38 | 2.32 |
| 100 | 5.18 | 3.83 | 3.25 | 2.92 | 2.70 | 2.54 | 2.42 | 2.32 | 2.24 | 2.18 |

(*Contd.*)

| $v_1 \backslash v_2$ | *11* | *12* | *15* | *20* | *25* | *30* | *40* | *50* | *100* |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 973.0 | 976.7 | 984.9 | 993.1 | 998.1 | 1001 | 1006 | 1008 | 1013 |
| 2 | 39.41 | 39.41 | 39.43 | 39.45 | 39.46 | 39.46 | 39.47 | 39.48 | 39.49 |
| 3 | 14.37 | 14.34 | 14.25 | 14.17 | 14.12 | 14.08 | 14.04 | 14.01 | 13.96 |
| 4 | 8.79 | 8.75 | 8.66 | 8.56 | 8.50 | 8.46 | 8.41 | 8.38 | 8.32 |
| 5 | 6.57 | 6.52 | 6.43 | 6.33 | 6.27 | 6.23 | 6.18 | 6.14 | 6.08 |
| 6 | 5.41 | 5.37 | 5.27 | 5.17 | 5.11 | 5.07 | 5.01 | 4.98 | 4.92 |
| 7 | 4.71 | 4.67 | 4.57 | 4.47 | 4.40 | 4.36 | 4.31 | 4.28 | 4.21 |
| 8 | 4.24 | 4.20 | 4.10 | 4.00 | 3.94 | 3.89 | 3.84 | 3.81 | 3.74 |
| 9 | 3.91 | 3.87 | 3.77 | 3.67 | 3.60 | 3.56 | 3.51 | 3.47 | 3.40 |
| 10 | 3.66 | 3.62 | 3.52 | 3.42 | 3.35 | 3.31 | 3.26 | 3.22 | 3.15 |
| 11 | 3.47 | 3.43 | 3.33 | 3.23 | 3.16 | 3.12 | 3.06 | 3.03 | 2.96 |
| 12 | 3.32 | 3.28 | 3.18 | 3.07 | 3.01 | 2.96 | 2.91 | 2.87 | 2.80 |
| 13 | 3.20 | 3.15 | 3.05 | 2.95 | 2.88 | 2.84 | 2.78 | 2.74 | 2.67 |
| 14 | 3.09 | 3.05 | 2.95 | 2.84 | 2.78 | 2.73 | 2.67 | 2.64 | 2.56 |
| 15 | 3.01 | 2.96 | 2.86 | 2.76 | 2.69 | 2.64 | 2.59 | 2.55 | 2.47 |
| 16 | 2.93 | 2.89 | 2.79 | 2.68 | 2.61 | 2.57 | 2.51 | 2.47 | 2.40 |
| 17 | 2.87 | 2.82 | 2.72 | 2.62 | 2.55 | 2.50 | 2.44 | 2.41 | 2.33 |
| 18 | 2.81 | 2.77 | 2.67 | 2.56 | 2.49 | 2.44 | 2.38 | 2.35 | 2.27 |
| 19 | 2.76 | 2.72 | 2.62 | 2.51 | 2.44 | 2.39 | 2.33 | 2.30 | 2.22 |
| 20 | 2.72 | 2.68 | 2.57 | 2.46 | 2.40 | 2.35 | 2.29 | 2.25 | 2.17 |
| 21 | 2.68 | 2.64 | 2.53 | 2.42 | 2.36 | 2.31 | 2.25 | 2.21 | 2.13 |
| 22 | 2.65 | 2.60 | 2.50 | 2.39 | 2.32 | 2.27 | 2.21 | 2.17 | 2.09 |
| 23 | 2.62 | 2.57 | 2.47 | 2.36 | 2.29 | 2.24 | 2.18 | 2.14 | 2.06 |
| 24 | 2.59 | 2.54 | 2.44 | 2.33 | 2.26 | 2.21 | 2.15 | 2.11 | 2.02 |
| 25 | 2.56 | 2.51 | 2.41 | 2.30 | 2.23 | 2.18 | 2.12 | 2.08 | 2.00 |
| 30 | 2.46 | 2.41 | 2.31 | 2.20 | 2.12 | 2.07 | 2.01 | 1.97 | 1.88 |
| 40 | 2.33 | 2.29 | 2.18 | 2.07 | 1.99 | 1.94 | 1.88 | 1.83 | 1.74 |
| 50 | 2.26 | 2.22 | 2.11 | 1.99 | 1.92 | 1.87 | 1.80 | 1.75 | 1.66 |
| 100 | 2.12 | 2.08 | 1.97 | 1.85 | 1.77 | 1.71 | 1.64 | 1.59 | 1.48 |

(*Contd.*)

## A.3 PERCENTAGE POINTS OF THE *F* DISTRIBUTION: $F_{0.01,\nu_1,\nu_2}$ ($\nu_1$ = Numerator *df* and $\nu_2$ = denominator *df*)

| $\nu_1 \backslash \nu_2$ | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 4052 | 5000 | 5403 | 5625 | 5764 | 5859 | 5928 | 5981 | 6022 | 6056 |
| 2 | 98.50 | 99.00 | 99.17 | 99.25 | 99.30 | 99.33 | 99.36 | 99.37 | 99.39 | 99.40 |
| 3 | 34.12 | 30.82 | 29.46 | 28.71 | 28.24 | 27.91 | 27.67 | 27.49 | 27.35 | 27.23 |
| 4 | 21.20 | 18.00 | 16.69 | 15.98 | 15.52 | 15.21 | 14.98 | 14.80 | 14.66 | 14.55 |
| 5 | 16.26 | 13.27 | 12.06 | 11.39 | 10.97 | 10.67 | 10.46 | 10.29 | 10.16 | 10.05 |
| 6 | 13.75 | 10.92 | 9.78 | 9.15 | 8.75 | 8.47 | 8.26 | 8.10 | 7.98 | 7.87 |
| 7 | 12.25 | 9.55 | 8.45 | 7.85 | 7.46 | 7.19 | 6.99 | 6.84 | 6.72 | 6.62 |
| 8 | 11.26 | 8.65 | 7.59 | 7.01 | 6.63 | 6.37 | 6.18 | 6.03 | 5.91 | 5.81 |
| 9 | 10.56 | 8.02 | 6.99 | 6.42 | 6.06 | 5.80 | 5.61 | 5.47 | 5.35 | 5.26 |
| 10 | 10.04 | 7.56 | 6.55 | 5.99 | 5.64 | 5.39 | 5.20 | 5.06 | 4.94 | 4.85 |
| 11 | 9.65 | 7.21 | 6.22 | 5.67 | 5.32 | 5.07 | 4.89 | 4.74 | 4.63 | 4.54 |
| 12 | 9.33 | 6.93 | 5.95 | 5.41 | 5.06 | 4.82 | 4.64 | 4.50 | 4.39 | 4.30 |
| 13 | 9.07 | 6.70 | 5.74 | 5.21 | 4.86 | 4.62 | 4.44 | 4.30 | 4.19 | 4.10 |
| 14 | 8.86 | 6.51 | 5.56 | 5.04 | 4.69 | 4.46 | 4.28 | 4.14 | 4.03 | 3.94 |
| 15 | 8.68 | 6.36 | 5.42 | 4.89 | 4.56 | 4.32 | 4.14 | 4.00 | 3.89 | 3.80 |
| 16 | 8.53 | 6.23 | 5.29 | 4.77 | 4.44 | 4.20 | 4.03 | 3.89 | 3.78 | 3.69 |
| 17 | 8.40 | 6.11 | 5.18 | 4.67 | 4.34 | 4.10 | 3.93 | 3.79 | 3.68 | 3.59 |
| 18 | 8.29 | 6.01 | 5.09 | 4.58 | 4.25 | 4.01 | 3.84 | 3.71 | 3.60 | 3.51 |
| 19 | 8.18 | 5.93 | 5.01 | 4.50 | 4.17 | 3.94 | 3.77 | 3.63 | 3.52 | 3.43 |
| 20 | 8.10 | 5.85 | 4.94 | 4.43 | 4.10 | 3.87 | 3.70 | 3.56 | 3.46 | 3.37 |
| 21 | 8.02 | 5.78 | 4.87 | 4.37 | 4.04 | 3.81 | 3.64 | 3.51 | 3.40 | 3.31 |
| 22 | 7.95 | 5.72 | 4.82 | 4.31 | 3.99 | 3.76 | 3.59 | 3.45 | 3.35 | 3.26 |
| 23 | 7.88 | 5.66 | 4.76 | 4.26 | 3.94 | 3.71 | 3.54 | 3.41 | 3.30 | 3.21 |
| 24 | 7.82 | 5.61 | 4.72 | 4.22 | 3.90 | 3.67 | 3.50 | 3.36 | 3.26 | 3.17 |
| 25 | 7.77 | 5.57 | 4.68 | 4.18 | 3.85 | 3.63 | 3.46 | 3.32 | 3.22 | 3.13 |
| 30 | 7.56 | 5.39 | 4.51 | 4.02 | 3.70 | 3.47 | 3.30 | 3.17 | 3.07 | 2.98 |
| 40 | 7.31 | 5.18 | 4.31 | 3.83 | 3.51 | 3.29 | 3.12 | 2.99 | 2.89 | 2.80 |
| 50 | 7.17 | 5.06 | 4.20 | 3.72 | 3.41 | 3.19 | 3.02 | 2.89 | 2.78 | 2.70 |
| 100 | 6.90 | 4.82 | 3.98 | 3.51 | 3.21 | 2.99 | 2.82 | 2.69 | 2.59 | 2.50 |

(*Contd.*)

| $v_1 \backslash v_2$ | *11* | *12* | *15* | *20* | *25* | *30* | *40* | *50* | *100* |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 6083 | 6106 | 6157 | 6209 | 6240 | 6261 | 6287 | 6303 | 6334 |
| 2 | 99.41 | 99.42 | 99.43 | 99.45 | 99.46 | 99.47 | 99.47 | 99.48 | 99.49 |
| 3 | 27.13 | 27.05 | 26.87 | 26.69 | 26.58 | 26.50 | 26.41 | 26.35 | 26.24 |
| 4 | 14.45 | 14.37 | 14.20 | 14.02 | 13.91 | 13.84 | 13.75 | 13.69 | 13.58 |
| 5 | 9.96 | 9.89 | 9.72 | 9.55 | 9.45 | 9.38 | 9.29 | 9.24 | 9.13 |
| 6 | 7.79 | 7.72 | 7.56 | 7.40 | 7.30 | 7.23 | 7.14 | 7.09 | 6.99 |
| 7 | 6.54 | 6.47 | 6.31 | 6.16 | 6.06 | 5.99 | 5.91 | 5.86 | 5.75 |
| 8 | 5.73 | 5.67 | 5.52 | 5.36 | 5.26 | 5.20 | 5.12 | 5.07 | 4.96 |
| 9 | 5.18 | 5.11 | 4.96 | 4.81 | 4.71 | 4.65 | 4.57 | 4.52 | 4.41 |
| 10 | 4.77 | 4.71 | 4.56 | 4.41 | 4.31 | 4.25 | 4.17 | 4.12 | 4.01 |
| 11 | 4.46 | 4.40 | 4.25 | 4.10 | 4.01 | 3.94 | 3.86 | 3.81 | 3.71 |
| 12 | 4.22 | 4.16 | 4.01 | 3.86 | 3.76 | 3.70 | 3.62 | 3.57 | 3.47 |
| 13 | 4.02 | 3.96 | 3.82 | 3.66 | 3.57 | 3.51 | 3.43 | 3.38 | 3.27 |
| 14 | 3.86 | 3.80 | 3.66 | 3.51 | 3.41 | 3.35 | 3.27 | 3.22 | 3.11 |
| 15 | 3.73 | 3.67 | 3.52 | 3.37 | 3.28 | 3.21 | 3.13 | 3.08 | 2.98 |
| 16 | 3.62 | 3.55 | 3.41 | 3.26 | 3.16 | 3.10 | 3.02 | 2.97 | 2.86 |
| 17 | 3.52 | 3.46 | 3.31 | 3.16 | 3.07 | 3.00 | 2.92 | 2.87 | 2.76 |
| 18 | 3.43 | 3.37 | 3.23 | 3.08 | 2.98 | 2.92 | 2.84 | 2.78 | 2.68 |
| 19 | 3.36 | 3.30 | 3.15 | 3.00 | 2.91 | 2.84 | 2.76 | 2.71 | 2.60 |
| 20 | 3.29 | 3.23 | 3.09 | 2.94 | 2.84 | 2.78 | 2.69 | 2.64 | 2.54 |
| 21 | 3.24 | 3.17 | 3.03 | 2.88 | 2.79 | 2.72 | 2.64 | 2.58 | 2.48 |
| 22 | 3.18 | 3.12 | 2.98 | 2.83 | 2.73 | 2.67 | 2.58 | 2.53 | 2.42 |
| 23 | 3.14 | 3.07 | 2.93 | 2.78 | 2.69 | 2.62 | 2.54 | 2.48 | 2.37 |
| 24 | 3.09 | 3.03 | 2.89 | 2.74 | 2.64 | 2.58 | 2.49 | 2.44 | 2.33 |
| 25 | 3.06 | 2.99 | 2.85 | 2.70 | 2.60 | 2.54 | 2.45 | 2.40 | 2.29 |
| 30 | 2.91 | 2.84 | 2.70 | 2.55 | 2.45 | 2.39 | 2.30 | 2.25 | 2.13 |
| 40 | 2.73 | 2.66 | 2.52 | 2.37 | 2.27 | 2.20 | 2.11 | 2.06 | 1.94 |
| 50 | 2.63 | 2.56 | 2.42 | 2.27 | 2.17 | 2.10 | 2.01 | 1.95 | 1.82 |
| 100 | 2.43 | 2.37 | 2.22 | 2.07 | 1.97 | 1.89 | 1.80 | 1.74 | 1.60 |

## A.4 PERCENTAGE POINTS OF STUDENTIZED RANGE STATISTIC $q_{0.05}(p, f)$

| | $p$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $f$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 18.1 | 26.7 | 32.8 | 37.2 | 40.5 | 43.1 | 45.4 | 47.3 | 49.1 |
| 2 | 6.09 | 8.28 | 9.80 | 10.89 | 11.73 | 12.43 | 13.03 | 13.54 | 13.99 |
| 3 | 4.50 | 5.88 | 6.83 | 7.51 | 8.04 | 8.47 | 8.85 | 9.18 | 9.46 |
| 4 | 3.93 | 5.00 | 5.76 | 6.31 | 6.73 | 7.06 | 7.35 | 7.60 | 7.83 |
| 5 | 3.64 | 4.60 | 5.22 | 5.67 | 6.03 | 6.33 | 6.58 | 6.80 | 6.99 |
| 6 | 3.46 | 4.34 | 4.90 | 5.31 | 5.63 | 5.89 | 6.12 | 6.32 | 6.49 |
| 7 | 3.34 | 4.16 | 4.68 | 5.06 | 5.35 | 5.59 | 5.80 | 5.99 | 6.15 |
| 8 | 3.26 | 4.04 | 4.53 | 4.89 | 5.17 | 5.40 | 5.60 | 5.77 | 5.92 |
| 9 | 3.20 | 3.95 | 4.42 | 4.76 | 5.02 | 5.24 | 5.43 | 5.60 | 5.74 |
| 10 | 3.15 | 3.88 | 4.33 | 4.66 | 4.91 | 5.12 | 5.30 | 5.46 | 5.60 |
| 11 | 3.11 | 3.82 | 4.26 | 4.58 | 4.82 | 5.03 | 5.20 | 5.35 | 5.49 |
| 12 | 3.08 | 3.77 | 4.20 | 4.51 | 4.75 | 4.95 | 5.12 | 5.27 | 5.40 |
| 13 | 3.06 | 3.73 | 4.15 | 4.46 | 4.69 | 4.88 | 5.05 | 5.19 | 5.32 |
| 14 | 3.03 | 3.70 | 4.11 | 4.41 | 4.64 | 4.83 | 4.99 | 5.13 | 5.25 |
| 15 | 3.01 | 3.67 | 4.08 | 4.37 | 4.59 | 4.78 | 4.94 | 5.08 | 5.20 |
| 16 | 3.00 | 3.65 | 4.05 | 4.34 | 4.56 | 4.74 | 4.90 | 5.03 | 5.15 |
| 17 | 2.98 | 3.62 | 4.02 | 4.31 | 4.52 | 4.70 | 4.86 | 4.99 | 5.11 |
| 18 | 2.97 | 3.61 | 4.00 | 4.28 | 4.49 | 4.67 | 4.83 | 4.96 | 5.07 |
| 19 | 2.96 | 3.59 | 3.98 | 4.26 | 4.47 | 4.64 | 4.79 | 4.92 | 5.04 |
| 20 | 2.95 | 3.58 | 3.96 | 4.24 | 4.45 | 4.62 | 4.77 | 4.90 | 5.01 |
| 24 | 2.92 | 3.53 | 3.90 | 4.17 | 4.37 | 4.54 | 4.68 | 4.81 | 4.92 |
| 30 | 2.89 | 3.48 | 3.84 | 4.11 | 4.30 | 4.46 | 4.60 | 4.72 | 4.83 |
| 40 | 2.86 | 3.44 | 3.79 | 4.04 | 4.23 | 4.39 | 4.52 | 4.63 | 4.74 |
| 60 | 2.83 | 3.40 | 3.74 | 3.98 | 4.16 | 4.31 | 4.44 | 4.55 | 4.65 |
| 120 | 2.80 | 3.36 | 3.69 | 3.92 | 4.10 | 4.24 | 4.36 | 4.47 | 4.56 |
| $\infty$ | 2.77 | 3.32 | 3.63 | 3.86 | 4.03 | 4.17 | 4.29 | 4.39 | 4.47 |

## A.4 PERCENTAGE POINTS OF STUDENTIZED RANGE STATISTIC $q_{0.01}(p, f)$

| | p | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $f$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 90.0 | 135 | 164 | 186 | 202 | 216 | 227 | 237 | 246 |
| 2 | 14.0 | 19.0 | 22.3 | 24.7 | 26.6 | 28.2 | 29.5 | 30.7 | 31.7 |
| 3 | 8.26 | 10.6 | 12.2 | 13.3 | 14.2 | 15.0 | 15.6 | 16.2 | 16.7 |
| 4 | 6.51 | 8.12 | 9.17 | 9.96 | 10.6 | 11.1 | 11.5 | 11.9 | 12.3 |
| 5 | 5.70 | 6.97 | 7.80 | 8.42 | 8.91 | 9.32 | 9.67 | 9.97 | 10.24 |
| 6 | 5.24 | 6.33 | 7.03 | 7.56 | 7.97 | 8.32 | 8.61 | 8.87 | 9.10 |
| 7 | 4.95 | 5.92 | 6.54 | 7.01 | 7.37 | 7.68 | 7.94 | 8.17 | 8.37 |
| 8 | 4.74 | 5.63 | 6.20 | 6.63 | 6.96 | 7.24 | 7.47 | 7.68 | 7.87 |
| 9 | 4.60 | 5.43 | 5.96 | 6.35 | 6.66 | 6.91 | 7.13 | 7.32 | 7.49 |
| 10 | 4.48 | 5.27 | 5.77 | 6.14 | 6.43 | 6.67 | 6.87 | 7.05 | 7.21 |
| 11 | 4.39 | 5.14 | 5.62 | 5.97 | 6.25 | 6.48 | 6.67 | 6.84 | 6.99 |
| 12 | 4.32 | 5.04 | 5.50 | 5.84 | 6.10 | 6.32 | 6.51 | 6.67 | 6.81 |
| 13 | 4.26 | 4.96 | 5.40 | 5.73 | 5.98 | 6.19 | 6.37 | 6.53 | 6.67 |
| 14 | 4.21 | 4.89 | 5.32 | 5.63 | 5.88 | 6.08 | 6.26 | 6.41 | 6.54 |
| 15 | 4.17 | 4.83 | 5.25 | 5.56 | 5.80 | 5.99 | 6.16 | 6.31 | 6.44 |
| 16 | 4.13 | 4.78 | 5.19 | 5.49 | 5.72 | 5.92 | 6.08 | 6.22 | 6.35 |
| 17 | 4.10 | 4.74 | 5.14 | 5.43 | 5.66 | 5.85 | 6.01 | 6.15 | 6.27 |
| 18 | 4.07 | 4.70 | 5.09 | 5.38 | 5.60 | 5.79 | 5.94 | 6.08 | 6.20 |
| 19 | 4.05 | 4.67 | 5.05 | 5.33 | 5.55 | 5.73 | 5.89 | 6.02 | 6.14 |
| 20 | 4.02 | 4.64 | 5.02 | 5.29 | 5.51 | 5.69 | 5.84 | 5.97 | 6.09 |
| 24 | 3.96 | 4.54 | 4.91 | 5.17 | 5.37 | 5.54 | 5.69 | 5.81 | 5.92 |
| 30 | 3.89 | 4.45 | 4.80 | 5.05 | 5.24 | 5.40 | 5.54 | 5.65 | 5.76 |
| 40 | 3.82 | 4.37 | 4.70 | 4.93 | 5.11 | 5.27 | 5.39 | 5.50 | 5.60 |
| 60 | 3.76 | 4.28 | 4.60 | 4.82 | 4.99 | 5.13 | 5.25 | 5.36 | 5.45 |
| 120 | 3.70 | 4.20 | 4.50 | 4.71 | 4.87 | 5.01 | 5.12 | 5.21 | 5.30 |
| $\infty$ | 3.64 | 4.12 | 4.40 | 4.60 | 4.76 | 4.88 | 4.99 | 5.08 | 5.16 |

## A.5 SIGNIFICANT RANGES FOR DUNCAN'S MULTIPLE RANGE TEST $r_{0.05}(p, f)$

| | *p* | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *f* | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 20 | 50 | 100 |
| 1 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 |
| 2 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 |
| 3 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 |
| 4 | 3.93 | 4.01 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 |
| 5 | 3.64 | 3.74 | 3.79 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 |
| 6 | 3.46 | 3.58 | 3.64 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 |
| 7 | 3.35 | 3.47 | 3.54 | 3.58 | 3.60 | 3.61 | 3.61 | 3.61 | 3.61 | 3.61 | 3.61 | 3.61 |
| 8 | 3.26 | 3.39 | 3.47 | 3.52 | 3.55 | 3.56 | 3.56 | 3.56 | 3.56 | 3.56 | 3.56 | 3.56 |
| 9 | 3.20 | 3.34 | 3.41 | 3.47 | 3.50 | 3.52 | 3.52 | 3.52 | 3.52 | 3.52 | 3.52 | 3.52 |
| 10 | 3.15 | 3.30 | 3.37 | 3.43 | 3.46 | 3.47 | 3.47 | 3.47 | 3.47 | 3.48 | 3.48 | 3.48 |
| 11 | 3.11 | 3.27 | 3.35 | 3.39 | 3.43 | 3.44 | 3.45 | 3.46 | 3.46 | 3.48 | 3.48 | 3.48 |
| 12 | 3.08 | 3.23 | 3.33 | 3.36 | 3.40 | 3.42 | 3.44 | 3.44 | 3.46 | 3.48 | 3.48 | 3.48 |
| 13 | 3.06 | 3.21 | 3.30 | 3.35 | 3.38 | 3.41 | 3.42 | 3.44 | 3.45 | 3.47 | 3.47 | 3.47 |
| 14 | 3.03 | 3.18 | 3.27 | 3.33 | 3.37 | 3.39 | 3.41 | 3.42 | 3.44 | 3.47 | 3.47 | 3.47 |
| 15 | 3.01 | 3.16 | 3.25 | 3.31 | 3.36 | 3.38 | 3.40 | 3.42 | 3.43 | 3.47 | 3.47 | 3.47 |
| 16 | 3.00 | 3.15 | 3.23 | 3.30 | 3.34 | 3.37 | 3.39 | 3.41 | 3.43 | 3.47 | 3.47 | 3.47 |
| 17 | 2.98 | 3.13 | 3.22 | 3.28 | 3.33 | 3.36 | 3.38 | 3.40 | 3.42 | 3.47 | 3.47 | 3.47 |
| 18 | 2.97 | 3.12 | 3.21 | 3.27 | 3.32 | 3.35 | 3.37 | 3.39 | 3.41 | 3.47 | 3.47 | 3.47 |
| 19 | 2.96 | 3.11 | 3.19 | 3.26 | 3.31 | 3.35 | 3.37 | 3.39 | 3.41 | 3.47 | 3.47 | 3.47 |
| 20 | 2.95 | 3.10 | 3.18 | 3.25 | 3.30 | 3.34 | 3.36 | 3.38 | 3.40 | 3.47 | 3.47 | 3.47 |
| 30 | 2.89 | 3.04 | 3.12 | 3.20 | 3.25 | 3.29 | 3.32 | 3.35 | 3.37 | 3.47 | 3.47 | 3.47 |
| 40 | 2.86 | 3.01 | 3.10 | 3.17 | 3.22 | 3.27 | 3.30 | 3.33 | 3.35 | 3.47 | 3.47 | 3.47 |
| 60 | 2.83 | 2.98 | 3.08 | 3.14 | 3.20 | 3.24 | 3.28 | 3.31 | 3.33 | 3.47 | 3.48 | 3.48 |
| 100 | 2.80 | 2.95 | 3.05 | 3.12 | 3.18 | 3.22 | 3.26 | 3.29 | 3.32 | 3.47 | 3.53 | 3.53 |
| ∞ | 2.77 | 2.92 | 3.02 | 3.09 | 3.15 | 3.19 | 3.23 | 3.26 | 3.29 | 3.47 | 3.61 | 3.67 |

## A.5 SIGNIFICANT RANGES FOR DUNCAN'S MULTIPLE RANGE TEST $r_{0.01}(p, f)$

| | $p$ | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $f$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 20 | 50 | 100 |
| 1 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 |
| 2 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 |
| 3 | 8.26 | 8.5 | 8.6 | 8.7 | 8.8 | 8.9 | 8.9 | 9.0 | 9.0 | 9.3 | 9.3 | 9.3 |
| 4 | 6.51 | 6.8 | 6.9 | 7.0 | 7.1 | 7.1 | 7.2 | 7.2 | 7.3 | 7.5 | 7.5 | 7.5 |
| 5 | 5.70 | 5.96 | 6.11 | 6.18 | 6.26 | 6.33 | 6.40 | 6.44 | 6.5 | 6.8 | 6.8 | 6.8 |
| 6 | 5.24 | 5.51 | 5.65 | 5.73 | 5.81 | 5.88 | 5.95 | 6.00 | 6.0 | 6.3 | 6.3 | 6.3 |
| 7 | 4.95 | 5.22 | 5.37 | 5.45 | 5.53 | 5.61 | 5.69 | 5.73 | 5.8 | 6.0 | 6.0 | 6.0 |
| 8 | 4.74 | 5.00 | 5.14 | 5.23 | 5.32 | 5.40 | 5.47 | 5.51 | 5.5 | 5.8 | 5.8 | 5.8 |
| 9 | 4.60 | 4.86 | 4.99 | 5.08 | 5.17 | 5.25 | 5.32 | 5.36 | 5.4 | 5.7 | 5.7 | 5.7 |
| 10 | 4.48 | 4.73 | 4.88 | 4.96 | 5.06 | 5.13 | 5.20 | 5.24 | 5.28 | 5.55 | 5.55 | 5.55 |
| 11 | 4.39 | 4.63 | 4.77 | 4.86 | 4.94 | 5.01 | 5.06 | 5.12 | 5.15 | 5.39 | 5.39 | 5.39 |
| 12 | 4.32 | 4.55 | 4.68 | 4.76 | 4.84 | 4.92 | 4.96 | 5.02 | 5.07 | 5.26 | 5.26 | 5.26 |
| 13 | 4.26 | 4.48 | 4.62 | 4.69 | 4.74 | 4.84 | 4.88 | 4.94 | 4.98 | 5.15 | 5.15 | 5.15 |
| 14 | 4.21 | 4.42 | 4.55 | 4.63 | 4.70 | 4.78 | 4.83 | 4.87 | 4.91 | 5.07 | 5.07 | 5.07 |
| 15 | 4.17 | 4.37 | 4.50 | 4.58 | 4.64 | 4.72 | 4.77 | 4.81 | 4.84 | 5.00 | 5.00 | 5.00 |
| 16 | 4.13 | 4.34 | 4.45 | 4.54 | 4.60 | 4.67 | 4.72 | 4.76 | 4.79 | 4.94 | 4.94 | 4.94 |
| 17 | 4.10 | 4.30 | 4.41 | 4.50 | 4.56 | 4.63 | 4.68 | 4.73 | 4.75 | 4.89 | 4.89 | 4.89 |
| 18 | 4.07 | 4.27 | 4.38 | 4.46 | 4.53 | 4.59 | 4.64 | 4.68 | 4.71 | 4.85 | 4.85 | 4.85 |
| 19 | 4.05 | 4.24 | 4.35 | 4.43 | 4.50 | 4.56 | 4.61 | 4.64 | 4.67 | 4.82 | 4.82 | 4.82 |
| 20 | 4.02 | 4.22 | 4.33 | 4.40 | 4.47 | 4.53 | 4.58 | 4.61 | 4.65 | 4.79 | 4.79 | 4.79 |
| 30 | 3.89 | 4.06 | 4.16 | 4.22 | 4.32 | 4.36 | 4.41 | 4.45 | 4.48 | 4.65 | 4.71 | 4.71 |
| 40 | 3.82 | 3.99 | 4.10 | 4.17 | 4.24 | 4.30 | 4.34 | 4.37 | 4.41 | 4.59 | 4.69 | 4.69 |
| 60 | 3.76 | 3.92 | 4.03 | 4.12 | 4.17 | 4.23 | 4.27 | 4.31 | 4.34 | 4.53 | 4.66 | 4.66 |
| 100 | 3.71 | 3.86 | 3.98 | 4.06 | 4.11 | 4.17 | 4.21 | 4.25 | 4.29 | 4.48 | 4.64 | 4.65 |
| ∞ | 3.64 | 3.80 | 3.90 | 3.98 | 4.04 | 4.09 | 4.14 | 4.17 | 4.20 | 4.41 | 4.60 | 4.68 |

# APPENDIX B

## B.1 TWO-LEVEL ORTHOGONAL ARRAYS

**$L_4$ Standard Array**

| *Trial no.* | *Column no.* | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| 1 | 1 | 1 | 1 |
| 2 | 1 | 2 | 2 |
| 3 | 2 | 1 | 2 |
| 4 | 2 | 2 | 1 |

**$L_8$ Standard Array**

| *Trial no.* | *Column no.* | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 |

### $L_{12}$ Standard Array*

| Trial no. | Column no. | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* | *11* |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 |
| 4 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 2 |
| 5 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 |
| 6 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 |
| 7 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 |
| 8 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 |
| 9 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 1 |
| 10 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 2 |
| 11 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 2 | 2 |
| 12 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 |

*No specific interaction columns are available.

### $L_{16}$ Standard Array

| Trial no. | Column no. | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* | *11* | *12* | *13* | *14* | *15* |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 |
| 4 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 |
| 5 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 |
| 6 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 |
| 7 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 |
| 8 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 |
| 9 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| 10 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 |
| 11 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 |
| 12 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 |
| 13 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 |
| 14 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 |
| 15 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 |
| 16 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 |

**$L_{32}$ Standard Array**

| Trial no. | Column no. | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* | *11* | *12* | *13* | *14* | *15* | *16* | *17* | *18* |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 |
| 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 |
| 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 5 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 |
| 6 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 7 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 8 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 |
| 9 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 |
| 10 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 |
| 11 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 2 |
| 12 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 |
| 13 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 |
| 14 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 |
| 15 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 2 |
| 16 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 |
| 17 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 |
| 18 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 |
| 19 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 |
| 20 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| 21 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 |
| 22 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| 23 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 1 |
| 24 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 |
| 25 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 |
| 26 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 |
| 27 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 |
| 28 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 |
| 29 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 |
| 30 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 |
| 31 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 |
| 32 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 |

(*Contd.*)

***L*$_{32}$ Standard Array (*Contd.*)**

| *Trial no.* | *Column no.* | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *19* | *20* | *21* | *22* | *23* | *24* | *25* | *26* | *27* | *28* | *29* | *30* | *31* |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 4 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 5 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 |
| 6 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 |
| 7 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 |
| 8 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 |
| 9 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 |
| 10 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 |
| 11 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 |
| 12 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 |
| 13 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 |
| 14 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 |
| 15 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 |
| 16 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 |
| 17 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| 18 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 |
| 19 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 |
| 20 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| 21 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 |
| 22 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 |
| 23 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 |
| 24 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 |
| 25 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 |
| 26 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 |
| 27 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 |
| 28 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 |
| 29 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 |
| 30 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 |
| 31 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 |
| 32 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 |

**Two-level Interaction Table**

| Column no. | Column no. | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* | *11* | *12* | *13* | *14* | *15* |
| 1 | 3 | 2 | 5 | 4 | 7 | 6 | 9 | 8 | 11 | 10 | 13 | 12 | 15 | 14 |
| 2 | – | 1 | 6 | 7 | 4 | 5 | 10 | 11 | 8 | 9 | 14 | 15 | 12 | 13 |
| 3 | – | – | 7 | 6 | 5 | 4 | 11 | 10 | 9 | 8 | 15 | 14 | 13 | 12 |
| 4 | – | – | – | 1 | 2 | 3 | 12 | 13 | 14 | 15 | 8 | 9 | 10 | 11 |
| 5 | – | – | – | – | 3 | 2 | 13 | 12 | 15 | 14 | 9 | 8 | 11 | 10 |
| 6 | – | – | – | – | – | 1 | 14 | 15 | 12 | 13 | 10 | 11 | 8 | 9 |
| 7 | – | – | – | – | – | – | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 |
| 8 | – | – | – | – | – | – | – | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 9 | – | – | – | – | – | – | – | – | 3 | 2 | 5 | 4 | 7 | 6 |
| 10 | – | – | – | – | – | – | – | – | – | 1 | 6 | 7 | 4 | 5 |
| 11 | – | – | – | – | – | – | – | – | – | – | 7 | 6 | 5 | 4 |
| 12 | – | – | – | – | – | – | – | – | – | – | – | 1 | 2 | 3 |
| 13 | – | – | – | – | – | – | – | – | – | – | – | – | 3 | 2 |
| 14 | – | – | – | – | – | – | – | – | – | – | – | – | – | 1 |
| 15 | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| 16 | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| 17 | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| 18 | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| 19 | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| 20 | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| 21 | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| 22 | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| 23 | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| 24 | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| 25 | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| 26 | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| 27 | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| 28 | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| 29 | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| 30 | – | – | – | – | – | – | – | – | – | – | – | – | – | – |

(*Contd.*)

**Two-level Interaction Table**

| *Column* | *Column no.* | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *no.* | *16* | *17* | *18* | *19* | *20* | *21* | *22* | *23* | *24* | *25* | *26* | *27* | *28* | *29* | *30* | *31* |
| 1 | 17 | 16 | 19 | 18 | 21 | 20 | 23 | 22 | 25 | 24 | 27 | 26 | 29 | 28 | 31 | 30 |
| 2 | 18 | 19 | 16 | 17 | 22 | 23 | 20 | 21 | 26 | 27 | 24 | 25 | 30 | 31 | 28 | 29 |
| 3 | 19 | 18 | 17 | 16 | 23 | 22 | 21 | 20 | 27 | 26 | 25 | 24 | 31 | 30 | 29 | 28 |
| 4 | 20 | 21 | 22 | 23 | 16 | 17 | 18 | 19 | 28 | 29 | 30 | 31 | 24 | 25 | 26 | 27 |
| 5 | 21 | 20 | 23 | 22 | 17 | 16 | 19 | 18 | 29 | 28 | 31 | 30 | 25 | 24 | 27 | 26 |
| 6 | 22 | 23 | 20 | 21 | 18 | 19 | 16 | 17 | 30 | 31 | 28 | 29 | 26 | 27 | 24 | 25 |
| 7 | 23 | 22 | 21 | 20 | 19 | 18 | 17 | 16 | 31 | 30 | 29 | 28 | 27 | 26 | 25 | 24 |
| 8 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| 9 | 25 | 24 | 27 | 26 | 29 | 28 | 31 | 30 | 17 | 16 | 19 | 18 | 21 | 20 | 23 | 22 |
| 10 | 26 | 27 | 24 | 25 | 30 | 31 | 28 | 29 | 18 | 19 | 16 | 17 | 22 | 23 | 20 | 21 |
| 11 | 27 | 26 | 25 | 24 | 31 | 30 | 29 | 28 | 19 | 18 | 17 | 16 | 23 | 22 | 21 | 20 |
| 12 | 28 | 29 | 30 | 31 | 24 | 25 | 26 | 27 | 20 | 21 | 22 | 23 | 16 | 17 | 18 | 19 |
| 13 | 29 | 28 | 31 | 30 | 25 | 24 | 27 | 26 | 21 | 20 | 23 | 22 | 17 | 16 | 19 | 18 |
| 14 | 30 | 31 | 28 | 29 | 26 | 27 | 24 | 25 | 22 | 23 | 20 | 21 | 18 | 19 | 16 | 17 |
| 15 | 31 | 30 | 29 | 28 | 27 | 26 | 25 | 24 | 23 | 22 | 21 | 20 | 19 | 18 | 17 | 16 |
| 16 | – | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 17 | – | – | 3 | 2 | 5 | 4 | 7 | 6 | 9 | 8 | 11 | 10 | 13 | 12 | 15 | 14 |
| 18 | – | – | – | 1 | 6 | 7 | 4 | 5 | 10 | 11 | 8 | 9 | 14 | 15 | 12 | 13 |
| 19 | – | – | – | – | 7 | 6 | 5 | 4 | 11 | 10 | 9 | 8 | 15 | 14 | 13 | 12 |
| 20 | – | – | – | – | – | 1 | 2 | 3 | 12 | 13 | 14 | 15 | 8 | 9 | 10 | 11 |
| 21 | – | – | – | – | – | – | 3 | 2 | 13 | 12 | 15 | 14 | 9 | 8 | 11 | 10 |
| 22 | – | – | – | – | – | – | – | 1 | 14 | 15 | 12 | 13 | 10 | 11 | 8 | 9 |
| 23 | – | – | – | – | – | – | – | – | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 |
| 24 | – | – | – | – | – | – | – | – | – | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 25 | – | – | – | – | – | – | – | – | – | – | 3 | 2 | 5 | 4 | 7 | 6 |
| 26 | – | – | – | – | – | – | – | – | – | – | – | 1 | 6 | 7 | 4 | 5 |
| 27 | – | – | – | – | – | – | – | – | – | – | – | – | 7 | 6 | 5 | 4 |
| 28 | – | – | – | – | – | – | – | – | – | – | – | – | – | 1 | 2 | 3 |
| 29 | – | – | – | – | – | – | – | – | – | – | – | – | – | – | 3 | 2 |
| 30 | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | 1 |

## B.2 THREE-LEVEL ORTHOGONAL ARRAYS

**$L_9$ Standard Array**

| Trial no. | Column no. | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 2 | 2 | 2 |
| 3 | 1 | 3 | 3 | 3 |
| 4 | 2 | 1 | 2 | 3 |
| 5 | 2 | 2 | 3 | 1 |
| 6 | 2 | 3 | 1 | 2 |
| 7 | 3 | 1 | 3 | 2 |
| 8 | 3 | 2 | 1 | 3 |
| 9 | 3 | 3 | 2 | 1 |

**$L_{18}$ Standard Array***

| Trial | Column no. | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| no. | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 1 | 1 | 3 | 3 | 3 | 3 | 3 | 3 |
| 4 | 1 | 2 | 1 | 1 | 2 | 2 | 3 | 3 |
| 5 | 1 | 2 | 2 | 2 | 3 | 3 | 1 | 1 |
| 6 | 1 | 2 | 3 | 3 | 1 | 1 | 2 | 2 |
| 7 | 1 | 3 | 1 | 2 | 1 | 3 | 2 | 3 |
| 8 | 1 | 3 | 2 | 3 | 2 | 1 | 3 | 1 |
| 9 | 1 | 3 | 3 | 1 | 3 | 2 | 1 | 2 |
| 10 | 2 | 1 | 1 | 3 | 3 | 2 | 2 | 1 |
| 11 | 2 | 1 | 2 | 1 | 1 | 3 | 3 | 2 |
| 12 | 2 | 1 | 3 | 2 | 2 | 1 | 1 | 3 |
| 13 | 2 | 2 | 1 | 2 | 3 | 1 | 3 | 2 |
| 14 | 2 | 2 | 2 | 3 | 1 | 2 | 1 | 3 |
| 15 | 2 | 2 | 3 | 1 | 2 | 3 | 2 | 1 |
| 16 | 2 | 3 | 1 | 3 | 2 | 3 | 1 | 2 |
| 17 | 2 | 3 | 2 | 1 | 3 | 1 | 2 | 3 |
| 18 | 2 | 3 | 3 | 2 | 1 | 2 | 3 | 1 |

*Interaction between column 1 and 2 only allowed.

**$L_{27}$ Standard Array**

| *Trial no.* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 1 | 1 | 1 | 1 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 4 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 3 |
| 5 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 1 | 1 | 1 |
| 6 | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 1 | 1 | 1 | 2 | 2 | 2 |
| 7 | 1 | 3 | 3 | 3 | 1 | 1 | 1 | 3 | 3 | 3 | 2 | 2 | 2 |
| 8 | 1 | 3 | 3 | 3 | 2 | 2 | 2 | 1 | 1 | 1 | 3 | 3 | 3 |
| 9 | 1 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 1 | 1 | 1 |
| 10 | 2 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| 11 | 2 | 1 | 2 | 3 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 |
| 12 | 2 | 1 | 2 | 3 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 |
| 13 | 2 | 2 | 3 | 1 | 1 | 2 | 3 | 2 | 3 | 1 | 3 | 1 | 2 |
| 14 | 2 | 2 | 3 | 1 | 2 | 3 | 1 | 3 | 1 | 2 | 1 | 2 | 3 |
| 15 | 2 | 2 | 3 | 1 | 3 | 1 | 2 | 1 | 2 | 3 | 2 | 3 | 1 |
| 16 | 2 | 3 | 1 | 2 | 1 | 2 | 3 | 3 | 1 | 2 | 2 | 3 | 1 |
| 17 | 2 | 3 | 1 | 2 | 2 | 3 | 1 | 1 | 2 | 3 | 3 | 1 | 2 |
| 18 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 2 | 3 | 1 | 1 | 2 | 3 |
| 19 | 3 | 1 | 3 | 2 | 1 | 3 | 2 | 1 | 3 | 2 | 1 | 3 | 2 |
| 20 | 3 | 1 | 3 | 2 | 2 | 1 | 3 | 2 | 1 | 3 | 2 | 1 | 3 |
| 21 | 3 | 1 | 3 | 2 | 3 | 2 | 1 | 3 | 2 | 1 | 3 | 2 | 1 |
| 22 | 3 | 2 | 1 | 3 | 1 | 3 | 2 | 2 | 1 | 3 | 3 | 2 | 1 |
| 23 | 3 | 2 | 1 | 3 | 2 | 1 | 3 | 3 | 2 | 1 | 1 | 3 | 2 |
| 24 | 3 | 2 | 1 | 3 | 3 | 2 | 1 | 1 | 3 | 2 | 2 | 1 | 3 |
| 25 | 3 | 3 | 2 | 1 | 1 | 3 | 2 | 3 | 2 | 1 | 2 | 1 | 3 |
| 26 | 3 | 3 | 2 | 1 | 2 | 1 | 3 | 1 | 3 | 2 | 3 | 2 | 1 |
| 27 | 3 | 3 | 2 | 1 | 3 | 2 | 1 | 2 | 1 | 3 | 1 | 3 | 2 |

**Three-level Interaction Table (Does not apply to $L_{18}$)**

| *Column no.* | *Column no.* | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| 1 | 3 | 2 | 2 | 6 | 5 | 5 | 9 | 8 | 8 | 12 | 11 | 11 |
| 1 | 4 | 4 | 3 | 7 | 7 | 6 | 10 | 10 | 9 | 13 | 13 | 12 |
| 2 | – | 1 | 1 | 8 | 9 | 10 | 5 | 6 | 7 | 5 | 6 | 7 |
| 2 | – | 4 | 3 | 11 | 12 | 13 | 11 | 12 | 13 | 8 | 9 | 10 |
| 3 | – | – | 1 | 9 | 10 | 8 | 7 | 5 | 6 | 6 | 7 | 5 |
| 3 | – | – | 2 | 13 | 11 | 12 | 12 | 13 | 11 | 10 | 8 | 9 |
| 4 | – | – | – | 10 | 8 | 9 | 6 | 7 | 5 | 7 | 5 | 6 |
| 4 | – | – | – | 12 | 13 | 11 | 13 | 11 | 12 | 9 | 10 | 8 |
| 5 | – | – | – | – | 1 | 1 | 2 | 3 | 4 | 2 | 4 | 3 |
| 5 | – | – | – | – | 7 | 6 | 11 | 13 | 12 | 8 | 10 | 9 |
| 6 | – | – | – | – | – | 1 | 4 | 2 | 3 | 3 | 2 | 4 |
| 6 | – | – | – | – | – | 5 | 13 | 12 | 11 | 10 | 9 | 8 |
| 7 | – | – | – | – | – | – | 3 | 4 | 2 | 4 | 3 | 2 |
| 7 | – | – | – | – | – | – | 12 | 11 | 13 | 9 | 8 | 10 |
| 8 | – | – | – | – | – | – | – | 1 | 1 | 2 | 3 | 4 |
| 8 | – | – | – | – | – | – | – | 10 | 9 | 5 | 7 | 6 |
| 9 | – | – | – | – | – | – | – | – | 1 | 4 | 2 | 3 |
| 9 | – | – | – | – | – | – | – | – | 8 | 7 | 6 | 5 |
| 10 | – | – | – | – | – | – | – | – | – | 3 | 4 | 2 |
| 10 | – | – | – | – | – | – | – | – | – | 6 | 5 | 7 |
| 11 | – | – | – | – | – | – | – | – | – | – | 1 | 1 |
| 11 | – | – | – | – | – | – | – | – | – | – | 13 | 12 |
| 12 | – | – | – | – | – | – | – | – | – | – | – | 1 |
| 12 | – | – | – | – | – | – | – | – | – | – | – | 11 |

* *Source:* Taguchi and Konishi, Orthogonal arrays and Linear graphs: Tools for Quality Engineering, 1987, ASI Press.

# APPENDIX C

# Linear Graphs*

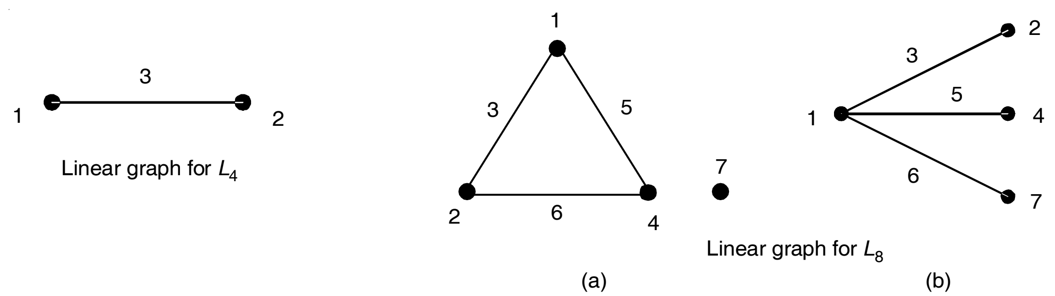



**FIGURE C.1** Standard linear graphs for $L_4$ and $L_8$ OAs.

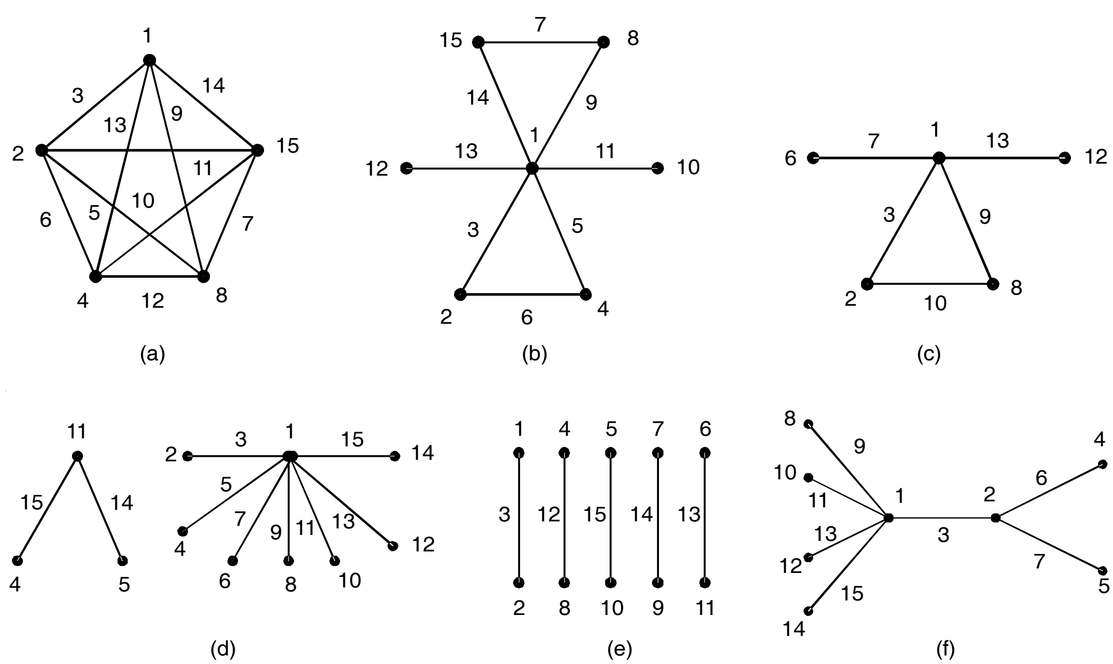



**FIGURE C.2** Standard linear graph for $L_{16}$ OA.

*Taguchi and Konishi, Orthogonal Arrays and Linear Graphs: Tools for Quality Engineering, 1987, ASI, Press.

# APPENDIX D

# Omega Conversion Table*

| P, % | db | P, % | db | P, % | db | P, % | db |
|---|---|---|---|---|---|---|---|
| 0.0 | ∞ | 3.0 | –15.0965 | 6.0 | –11.9498 | 9.0 | –10.0480 |
| 0.1 | –29.9957 | 3.1 | –14.9496 | 6.1 | –11.8734 | 9.1 | –9.9952 |
| 0.2 | –26.9810 | 3.2 | –14.8073 | 6.2 | –11.7981 | 9.2 | –9.9430 |
| 0.3 | –25.2157 | 3.3 | –14.6691 | 6.3 | –11.7240 | 9.3 | –9.8912 |
| 0.4 | –23.9620 | 3.4 | –14.5350 | 6.4 | –11.6510 | 9.4 | –9.8400 |
| 0.5 | –22.9885 | 3.5 | –14.4046 | 6.5 | –11.5790 | 9.5 | –9.7892 |
| 0.6 | –22.1924 | 3.6 | –14.2777 | 6.6 | –11.5080 | 9.6 | –9.7390 |
| 0.7 | –21.5185 | 3.7 | –14.1542 | 6.7 | –11.4381 | 9.7 | –9.6892 |
| 0.8 | –20.9342 | 3.8 | –14.0339 | 6.8 | –11.3691 | 9.8 | –9.6398 |
| 0.9 | –20.4183 | 3.9 | –13.9166 | 6.9 | –11.3010 | 9.9 | –9.5909 |
| 1.0 | –19.9564 | 4.0 | –13.8021 | 7.0 | –11.2338 | 10.0 | –9.5424 |
| 1.1 | –19.5380 | 4.1 | –13.6903 | 7.1 | –11.1676 | 10.1 | –9.4944 |
| 1.2 | –19.1558 | 4.2 | –13.5812 | 7.2 | –11.1022 | 10.2 | –9.4468 |
| 1.3 | –18.8037 | 4.3 | –13.4744 | 7.3 | –11.0376 | 10.3 | –9.3996 |
| 1.4 | –18.4775 | 4.4 | –13.3701 | 7.4 | –10.9738 | 10.4 | –9.3527 |
| 1.5 | –18.1734 | 4.5 | –13.2679 | 7.5 | –10.9108 | 10.5 | –9.3063 |
| 1.6 | –17.8888 | 4.6 | –13.1679 | 7.6 | –10.8486 | 10.6 | –9.2603 |
| 1.7 | –17.6210 | 4.7 | –13.0700 | 7.7 | –10.7871 | 10.7 | –9.2147 |
| 1.8 | –17.3684 | 4.8 | –12.9740 | 7.8 | –10.7264 | 10.8 | –9.1694 |
| 1.9 | –17.1292 | 4.9 | –12.8798 | 7.9 | –10.6663 | 10.9 | –9.1245 |
| 2.0 | –16.9020 | 5.0 | –12.7875 | 8.0 | –10.6070 | 11.0 | –9.0800 |
| 2.1 | –16.6856 | 5.1 | –12.6970 | 8.1 | –10.5483 | 11.1 | –9.0358 |
| 2.2 | –16.4792 | 5.2 | –12.6080 | 8.2 | –10.4903 | 11.2 | –8.9919 |
| 2.3 | –16.2817 | 5.3 | –12.5207 | 8.3 | –10.4329 | 11.3 | –8.9485 |
| 2.4 | –16.0924 | 5.4 | –12.4350 | 8.4 | –10.3762 | 11.4 | –8.9053 |
| 2.5 | –15.9106 | 5.5 | –12.3507 | 8.5 | –10.3200 | 11.5 | –8.8625 |
| 2.6 | –15.7359 | 5.6 | –12.2678 | 8.6 | –10.2645 | 11.6 | –8.8199 |
| 2.7 | –15.5675 | 5.7 | –12.1864 | 8.7 | –10.2095 | 11.7 | –8.7777 |
| 2.8 | –15.4051 | 5.8 | –12.1062 | 8.8 | –10.1551 | 11.8 | –8.7359 |
| 2.9 | –15.2482 | 5.9 | –12.0274 | 8.9 | –10.1013 | 11.9 | –8.6943 |

* $\Omega(\text{db}) = 10 \log \left[ \dfrac{p/100}{1 - p/100} \right]$

(*Contd.*)

| *P*, % | *db* | *P*, % | *db* | *P*, % | *db* | *P*, % | *db* |
|---|---|---|---|---|---|---|---|
| 12.0 | –8.6530 | 16.0 | –7.2016 | 20.0 | –6.0206 | 24.0 | –5.0060 |
| 12.1 | –8.6120 | 16.1 | –7.1694 | 20.1 | –5.9935 | 24.1 | –4.9822 |
| 12.2 | –8.5713 | 16.2 | –7.1373 | 20.2 | –5.9665 | 24.2 | –4.9585 |
| 12.3 | –8.5309 | 16.3 | –7.1054 | 20.3 | –5.9396 | 24.3 | –4.9349 |
| 12.4 | –8.4908 | 16.4 | –7.0736 | 20.4 | –5.9128 | 24.4 | –4.9113 |
| 12.5 | –8.4510 | 16.5 | –7.0420 | 20.5 | –5.8861 | 24.5 | –4.8878 |
| 12.6 | –8.4114 | 16.6 | –7.0106 | 20.6 | –5.8595 | 24.6 | –4.8644 |
| 12.7 | –8.3721 | 16.7 | –6.9793 | 20.7 | –5.8330 | 24.7 | –4.8410 |
| 12.8 | –8.3331 | 16.8 | –6.9481 | 20.8 | –5.8066 | 24.8 | –4.8177 |
| 12.9 | –8.2943 | 16.9 | –6.9171 | 20.9 | –5.7803 | 24.9 | –4.7944 |
| 13.0 | –8.2558 | 17.0 | –6.8863 | 21.0 | –5.7541 | 25.0 | –4.7712 |
| 13.1 | –8.2175 | 17.1 | –6.8556 | 21.1 | –5.7279 | 25.1 | –4.7481 |
| 13.2 | –8.1795 | 17.2 | –6.8250 | 21.2 | –5.7019 | 25.2 | –4.7250 |
| 13.3 | –8.1417 | 17.3 | –6.7946 | 21.3 | –5.6760 | 25.3 | –4.7020 |
| 13.4 | –8.1041 | 17.4 | –6.7643 | 21.4 | –5.6501 | 25.4 | –4.6791 |
| 13.5 | –8.0668 | 17.5 | –6.7342 | 21.5 | –5.6243 | 25.5 | –4.6562 |
| 13.6 | –8.0297 | 17.6 | –6.7041 | 21.6 | –5.5986 | 25.6 | –4.6333 |
| 13.7 | –7.9929 | 17.7 | –6.6743 | 21.7 | –5.5730 | 25.7 | –4.6106 |
| 13.8 | –7.9563 | 17.8 | –6.6445 | 21.8 | –5.5475 | 25.8 | –4.5878 |
| 13.9 | –7.9199 | 17.9 | –6.6149 | 21.9 | –5.5221 | 25.9 | –4.5652 |
| 14.0 | –7.8837 | 18.0 | –6.5854 | 22.0 | –5.4967 | 26.0 | –4.5426 |
| 14.1 | –7.8477 | 18.1 | –6.5561 | 22.1 | –5.4715 | 26.1 | –4.5200 |
| 14.2 | .–7.8120 | 18.2 | –6.5268 | 22.2 | –5.4463 | 26.2 | –4.4976 |
| 14.3 | –7.7764 | 18.3 | –6.4977 | 22.3 | –5.4212 | 26.3 | –4.4751 |
| 14.4 | –7.7411 | 18.4 | –6.4687 | 22.4 | –5.3961 | 26.4 | –4.4527 |
| 14.5 | –7.7060 | 18.5 | –6.4399 | 22.5 | –5.3712 | 26.5 | –4.4304 |
| 14.6 | –7.6711 | 18.6 | –6.4111 | 22.6 | –5.3463 | 26.6 | –4.4081 |
| 14.7 | –7.6363 | 18.7 | –6.3825 | 22.7 | –5.3215 | 26.7 | –4.3859 |
| 14.8 | –7.6018 | 18.8 | –6.3540 | 22.8 | –5.2968 | 26.8 | –4.3638 |
| 14.9 | –7.5674 | 18.9 | –6.3256 | 22.9 | –5.2722 | 26.9 | –4.3417 |
| 15.0 | –7.5333 | 19.0 | –6.2973 | 23.0 | –5.2476 | 27.0 | –4.3196 |
| 15.1 | –7.4993 | 19.1 | –6.2692 | 23.1 | –5.2231 | 27.1 | –4.2976 |
| 15.2 | –7.4655 | 19.2 | –6.2411 | 23.2 | –5.1987 | 27.2 | –4.2756 |
| 15.3 | –7.4319 | 19.3 | –6.2132 | 23.3 | –5.1744 | 27.3 | –4.2537 |
| 15.4 | –7.3985 | 19.4 | –6.1853 | 23.4 | –5.1501 | 27.4 | –4.2319 |
| 15.5 | –7.3653 | 19.5 | –6.1576 | 23.5 | –5.1259 | 27.5 | –4.2101 |
| 15.6 | –7.3322 | 19.6 | –6.1300 | 23.6 | –5.1018 | 27.6 | –4.1883 |
| 15.7 | –7.2993 | 19.7 | –6.1025 | 23.7 | –5.0778 | 27.7 | –4.1666 |
| 15.8 | –7.2666 | 19.8 | –6.0751 | 23.8 | –5.0538 | 27.8 | –4.1449 |
| 15.9 | –7.2340 | 19.9 | –6.0478 | 23.9 | –5.0299 | 27.9 | –4.1233 |

(*Contd.*)

| P, % | db | P, % | db | P, % | db | P, % | db |
|---|---|---|---|---|---|---|---|
| 28.0 | –4.1017 | 32.0 | –3.2736 | 36.0 | –2.4988 | 40.0 | –1.7609 |
| 28.1 | –4.0802 | 32.1 | –3.2536 | 36.1 | –2.4799 | 40.1 | –1.7428 |
| 28.2 | –4.0588 | 32.2 | –3.2337 | 36.2 | –2.4611 | 40.2 | –1.7248 |
| 28.3 | –4.0373 | 32.3 | –3.2139 | 36.3 | –2.4423 | 40.3 | –1.7067 |
| 28.4 | –4.0159 | 32.4 | –3.1940 | 36.4 | –2.4236 | 40.4 | –1.6886 |
| 28.5 | –3.9946 | 32.5 | –3.1742 | 36.5 | –2.4048 | 40.5 | –1.6706 |
| 28.6 | –3.9733 | 32.6 | –3.1544 | 36.6 | –2.3861 | 40.6 | –1.6526 |
| 28.7 | –3.9521 | 32.7 | –3.1347 | 36.7 | –2.3674 | 40.7 | –1.6346 |
| 28.8 | –3.9309 | 32.8 | –3.1150 | 36.8 | –2.3487 | 40.8 | –1.6166 |
| 28.9 | –3.9097 | 32.9 | –3.0953 | 36.9 | –2.3300 | 40.9 | –1.5986 |
| 29.0 | –3.8886 | 33.0 | –3.0756 | 37.0 | –2.3114 | 41.0 | –1.5807 |
| 29.1 | –3.8675 | 33.1 | –3.0560 | 37.1 | –2.2928 | 41.1 | –1.5627 |
| 29.2 | –3.8465 | 33.2 | –3.0364 | 37.2 | –2.2742 | 41.2 | –1.5448 |
| 29.3 | –3.8285 | 33.3 | –3.0168 | 37.3 | –2.2556 | 41.3 | –1.5269 |
| 29.4 | –3.8046 | 33.4 | –2.9973 | 37.4 | –2.2370 | 41.4 | –1.5090 |
| 29.5 | –3.7837 | 33.5 | –2.9778 | 37.5 | –2.2185 | 41.5 | –1.4911 |
| 29.6 | –3.7628 | 33.6 | –2.9583 | 37.6 | –2.2000 | 41.6 | –1.4732 |
| 29.7 | –3.7420 | 33.7 | –2.9388 | 37.7 | –2.1815 | 41.7 | –1.4553 |
| 29.8 | –3.7212 | 33.8 | –2.9194 | 37.8 | –2.1630 | 41.8 | –1.4375 |
| 29.9 | –3.7005 | 33.9 | –2.9000 | 37.9 | –2.1445 | 41.9 | –1.4196 |
| 30.0 | –3.6798 | 34.0 | –2.8807 | 38.0 | –2.1261 | 42.0 | –1.4018 |
| 30.1 | –3.6591 | 34.1 | –2.8613 | 38.1 | –2.1077 | 42.1 | –1.3840 |
| 30.2 | –3.6385 | 34.2 | –2.8420 | 38.2 | –2.0893 | 42.2 | –1.3662 |
| 30.3 | –3.6179 | 34.3 | –2.8227 | 38.3 | –2.0709 | 42.3 | –1.3484 |
| 30.4 | –3.5974 | 34.4 | –2.8035 | 38.4 | –2.0525 | 42.4 | –1.3306 |
| 30.5 | –3.5768 | 34.5 | –2.7842 | 38.5 | –2.0341 | 42.5 | –1.3128 |
| 30.6 | –3.5564 | 34.6 | –2.7650 | 38.6 | –2.0158 | 42.6 | –1.2950 |
| 30.7 | –3.5359 | 34.7 | –2.7458 | 38.7 | –1.9975 | 42.7 | –1.2773 |
| 30.8 | –3.5156 | 34.8 | –2.7267 | 38.8 | –1.9792 | 42.8 | –1.2595 |
| 30.9 | –3.4952 | 34.9 | –2.7076 | 38.9 | –1.9609 | 42.9 | –1.2418 |
| 31.0 | –3.4749 | 35.0 | –2.6885 | 39.0 | –1.9427 | 43.0 | –1.2241 |
| 31.1 | –3.4546 | 35.1 | –2.6694 | 39.1 | –1.9244 | 43.1 | –1.2063 |
| 31.2 | –3.4343 | 35.2 | –2.6503 | 39.2 | –1.9062 | 43.2 | –1.1886 |
| 31.3 | –3.4141 | 35.3 | –2.6313 | 39.3 | –1.8880 | 43.3 | –1.1710 |
| 31.4 | –3.3939 | 35.4 | –2.6123 | 39.4 | –1.8698 | 43.4 | –1.1533 |
| 31.5 | –3.3738 | 35.5 | –2.5933 | 39.5 | –1.8516 | 43.5 | –1.1356 |
| 31.6 | –3.3537 | 35.6 | –2.5744 | 39.6 | –1.8334 | 43.6 | –1.1179 |
| 31.7 | –3.3336 | 35.7 | –2.5554 | 39.7 | –1.8153 | 43.7 | –1.1003 |
| 31.8 | –3.3136 | 35.8 | –2.5365 | 39.8 | –1.7971 | 43.8 | –1.0826 |
| 31.9 | –3.2936 | 35.9 | –2.5176 | 39.9 | –1.7790 | 43.9 | –1.0650 |

(*Contd.*)

| *P*, % | *db* | *P*, % | *db* | *P*, % | *db* | *P*, % | *db* |
|---|---|---|---|---|---|---|---|
| 44.0 | –1.0474 | 48.0 | –0.3476 | 52.0 | 0.3476 | 56.0 | 1.0474 |
| 44.1 | –1.0297 | 48.1 | –0.3302 | 52.1 | 0.3650 | 56.1 | 1.0650 |
| 44.2 | –1.0121 | 48.2 | –0.3128 | 52.2 | 0.3824 | 56.2 | 1.0826 |
| 44.3 | –0.9945 | 48.3 | –0.2954 | 52.3 | 0.3998 | 56.3 | 1.1003 |
| 44.4 | –0.9769 | 48.4 | –0.2780 | 52.4 | 0.4172 | 56.4 | 1.1179 |
| 44.5 | –0.9593 | 48.5 | –0.2607 | 52.5 | 0.4347 | 56.5 | 1.1356 |
| 44.6 | –0.9417 | 48.6 | –0.2433 | 52.6 | 0.4521 | 56.6 | 1.1533 |
| 44.7 | –0.9242 | 48.7 | –0.2259 | 52.7 | 0.4695 | 56.7 | 1.1710 |
| 44.8 | –0.9066 | 48.8 | –0.2085 | 52.8 | 0.4869 | 56.8 | 1.1886 |
| 44.9 | –0.8891 | 48.9 | –0.1911 | 52.9 | 0.5043 | 56.9 | 1.2063 |
| 45.0 | –0.8715 | 49.0 | –0.1737 | 53.0 | 0.5218 | 57.0 | 1.2241 |
| 45.1 | –0.8540 | 49.1 | –0.1564 | 53.1 | 0.5392 | 57.1 | 1.2418 |
| 45.2 | –0.8364 | 49.2 | –0.1390 | 53.2 | 0.5567 | 57.2 | 1.2595 |
| 45.3 | –0.8189 | 49.3 | –0.1216 | 53.3 | 0.5741 | 57.3 | 1.2773 |
| 45.4 | –0.8014 | 49.4 | –0.1042 | 53.4 | 0.5916 | 57.4 | 1.2950 |
| 45.5 | –0.7839 | 49.5 | –0.0869 | 53.5 | 0.6090 | 57.5 | 1.3128 |
| 45.6 | –0.7663 | 49.6 | –0.0695 | 53.6 | 0.6265 | 57.6 | 1.3306 |
| 45.7 | –0.7488 | 49.7 | –0.0521 | 53.7 | 0.6439 | 57.7 | 1.3484 |
| 45.8 | –0.7313 | 49.8 | –0.0347 | 53.8 | 0.6614 | 57.8 | 1.3662 |
| 45.9 | –0.7138 | 49.9 | –0.0174 | 53.9 | 0.6789 | 57.9 | 1.3840 |
| 46.0 | –0.6964 | 50.0 | 0.0000 | 54.0 | 0.6964 | 58.0 | 1.4018 |
| 46.1 | –0.6789 | 50.1 | 0.0174 | 54.1 | 0.7138 | 58.1 | 1.4196 |
| 46.2 | –0.6614 | 50.2 | 0.0347 | 54.2 | 0.7313 | 58.2 | 1.4375 |
| 46.3 | –0.6439 | 50.3 | 0.0521 | 54.3 | 0.7488 | 58.3 | 1.4553 |
| 46.4 | –0.6265 | 50.4 | 0.0695 | 54.4 | 0.7663 | 58.4 | 1.4732 |
| 46.5 | –0.6090 | 50.5 | 0.0869 | 54.5 | 0.7839 | 58.5 | 1.4911 |
| 46.6 | –0.5916 | 50.6 | 0.1042 | 54.6 | 0.8014 | 58.6 | 1.5090 |
| 46.7 | –0.5741 | 50.7 | 0.1216 | 54.7 | 0.8189 | 58.7 | 1.5269 |
| 46.8 | –0.5567 | 50.8 | 0.1390 | 54.8 | 0.8364 | 58.8 | 1.5448 |
| 46.9 | –0.5392 | 50.9 | 0.1564 | 54.9 | 0.8540 | 58.9 | 1.5627 |
| 47.0 | –0.5218 | 51.0 | 0.1737 | 55.0 | 0.8715 | 59.0 | 1.5807 |
| 47.1 | –0.5043 | 51.1 | 0.1911 | 55.1 | 0.8891 | 59.1 | 1.5986 |
| 47.2 | –0.4869 | 51.2 | 0.2085 | 55.2 | 0.9066 | 59.2 | 1.6166 |
| 47.3 | –0.4695 | 51.3 | 0.2259 | 55.3 | 0.9242 | 59.3 | 1.6346 |
| 47.4 | –0.4521 | 51.4 | 0.2433 | 55.4 | 0.9417 | 59.4 | 1.6526 |
| 47.5 | –0.4347 | 51.5 | 0.2607 | 55.5 | 0.9593 | 59.5 | 1.6706 |
| 47.6 | –0.4172 | 51.6 | 0.2780 | 55.6 | 0.9769 | 59.6 | 1.6886 |
| 47.7 | –0.3998 | 51.7 | 0.2954 | 55.7 | 0.9945 | 59.7 | 1.7067 |
| 47.8 | –0.3824 | 51.8 | 0.1328 | 55.8 | 1.0121 | 59.8 | 1.7248 |
| 47.9 | –0.3650 | 51.9 | 0.3302 | 55.9 | 1.0297 | 59.9 | 1.7428 |

(*Contd.*)

| *P*, % | *db* | *P*, % | *db* | *P*, % | *db* | *P*, % | *db* |
|---|---|---|---|---|---|---|---|
| 60.0 | 1.7609 | 64.0 | 2.4988 | 68.0 | 3.2736 | 72.0 | 4.1017 |
| 60.1 | 1.7790 | 64.1 | 2.5176 | 68.1 | 3.2936 | 72.1 | 4.1233 |
| 60.2 | 1.7971 | 64.2 | 2.5365 | 68.2 | 3.3136 | 72.2 | 4.1449 |
| 60.3 | 1.8153 | 64.3 | 2.5554 | 68.3 | 3.3336 | 72.3 | 4.1666 |
| 60.4 | 1.8334 | 64.4 | 2.5744 | 68.4 | 3.3537 | 72.4 | 4.1883 |
| 60.5 | 1.8516 | 64.5 | 2.5933 | 68.5 | 3.3738 | 72.5 | 4.2101 |
| 60.6 | 1.8698 | 64.6 | 2.6123 | 68.6 | 3.3939 | 72.6 | 4.2319 |
| 60.7 | 1.8880 | 64.7 | 2.6313 | 68.7 | 3.4141 | 72.7 | 4.2537 |
| 60.8 | 1.9062 | 64.8 | 2.6503 | 68.8 | 3.4343 | 72.8 | 4.2756 |
| 60.9 | 1.9244 | 64.9 | 2.6694 | 68.9 | 3.4546 | 72.9 | 4.2976 |
| 61.0 | 1.9427 | 65.0 | 2.6885 | 69.0 | 3.4749 | 73.0 | 4.3196 |
| 61.1 | 1.9609 | 65.1 | 2.7076 | 69.1 | 3.4952 | 73.1 | 4.3417 |
| 61.2 | 1.9792 | 65.2 | 2.7267 | 69.2 | 3.5156 | 73.2 | 4.3638 |
| 61.3 | 1.9975 | 65.3 | 2.7458 | 69.3 | 3.5359 | 73.3 | 4.3859 |
| 61.4 | 2.0158 | 65.4 | 2.7650 | 69.4 | 3.5564 | 73.4 | 4.4081 |
| 61.5 | 2.0341 | 65.5 | 2.7842 | 69.5 | 3.5768 | 73.5 | 4.4304 |
| 61.6 | 2.0525 | 65.6 | 2.8035 | 69.6 | 3.5974 | 73.6 | 4.4527 |
| 61.7 | 2.0709 | 65.7 | 2.8227 | 69.7 | 3.6179 | 73.7 | 4.4751 |
| 61.8 | 2.0893 | 65.8 | 2.8420 | 69.8 | 3.6385 | 73.8 | 4.4976 |
| 61.9 | 2.1077 | 65.9 | 2.8613 | 69.9 | 3.6591 | 73.9 | 4.5200 |
| 62.0 | 2.1261 | 66.0 | 2.8807 | 70.0 | 3.6798 | 74.0 | 4.5426 |
| 62.1 | 2.1445 | 66.1 | 2.9000 | 70.1 | 3.7005 | 74.1 | 4.5652 |
| 62.2 | 2.1630 | 66.2 | 2.9194 | 70.2 | 3.7212 | 74.2 | 4.5878 |
| 62.3 | 2.1815 | 66.3 | 2.9388 | 70.3 | 3.7420 | 74.3 | 4.6106 |
| 62.4 | 2.2000 | 66.4 | 2.9583 | 70.4 | 3.7628 | 74.4 | 4.6333 |
| 62.5 | 2.2185 | 66.5 | 2.9778 | 70.5 | 3.7837 | 74.5 | 4.6562 |
| 62.6 | 2.2370 | 66.6 | 2.9973 | 70.6 | 3.8046 | 74.6 | 4.6791 |
| 62.7 | 2.2556 | 66.7 | 3.0168 | 70.7 | 3.8255 | 74.7 | 4.7020 |
| 62.8 | 2.2742 | 66.8 | 3.0364 | 70.8 | 3.8465 | 74.8 | 4.7250 |
| 62.9 | 2.2928 | 66.9 | 3.0560 | 70.9 | 3.8675 | 74.9 | 4.7481 |
| 63.0 | 2.3114 | 67.0 | 3.0756 | 71.0 | 3.8886 | 75.0 | 4.7712 |
| 63.1 | 2.3300 | 67.1 | 3.0953 | 71.1 | 3.9097 | 75.1 | 4.7944 |
| 63.2 | 2.3487 | 67.2 | 3.1150 | 71.2 | 3.9309 | 75.2 | 4.8177 |
| 63.3 | 2.3674 | 67.3 | 3.1347 | 71.3 | 3.9521 | 75.3 | 4.8410 |
| 63.4 | 2.3861 | 67.4 | 3.1544 | 71.4 | 3.9733 | 75.4 | 4.8644 |
| 63.5 | 2.4048 | 67.5 | 3.1742 | 71.5 | 3.9946 | 75.5 | 4.8878 |
| 63.6 | 2.4236 | 67.6 | 3.1940 | 71.6 | 4.0159 | 75.6 | 4.9113 |
| 63.7 | 2.4423 | 67.7 | 3.2139 | 71.7 | 4.0373 | 75.7 | 4.9349 |
| 63.8 | 2.4611 | 67.8 | 3.2337 | 71.8 | 4.0588 | 75.8 | 4.9585 |
| 63.9 | 2.4799 | 67.9 | 3.2536 | 71.9 | 4.0802 | 75.9 | 4.9822 |

(*Contd.*)

| *P*, % | *db* | *P*, % | *db* | *P*, % | *db* | *P*, % | *db* |
|---|---|---|---|---|---|---|---|
| 76.0 | 5.0060 | 80.0 | 6.0206 | 84.0 | 7.2016 | 88.0 | 8.6530 |
| 76.1 | 5.0299 | 80.1 | 6.0478 | 84.1 | 7.2340 | 88.1 | 8.6943 |
| 76.2 | 5.0538 | 80.2 | 6.0751 | 84.2 | 7.2666 | 88.2 | 8.7359 |
| 76.3 | 5.0778 | 80.3 | 6.1025 | 84.3 | 7.2993 | 88.3 | 8.7777 |
| 76.4 | 5.1018 | 80.4 | 6.1300 | 84.4 | 7.3322 | 88.4 | 8.8199 |
| 76.5 | 5.1259 | 80.5 | 6.1576 | 84.5 | 7.3653 | 88.5 | 8.8625 |
| 76.6 | 5.1501 | 80.6 | 6.1853 | 84.6 | 7.3985 | 88.6 | 8.9053 |
| 76.7 | 5.1744 | 80.7 | 6.2132 | 84.7 | 7.4319 | 88.7 | 8.9485 |
| 76.8 | 5.1987 | 80.8 | 6.2411 | 84.8 | 7.4655 | 88.8 | 8.9919 |
| 76.9 | 5.2231 | 80.9 | 6.2692 | 84.9 | 7.4993 | 88.9 | 9.0358 |
| 77.0 | 5.2476 | 81.0 | 6.2973 | 85.0 | 7.5333 | 89.0 | 9.0800 |
| 77.1 | 5.2722 | 81.1 | 6.3256 | 85.1 | 7.5674 | 89.1 | 9.1245 |
| 77.2 | 5.2968 | 81.2 | 6.3540 | 85.2 | 7.6018 | 89.2 | 9.1694 |
| 77.3 | 5.3215 | 81.3 | 6.3825 | 85.3 | 7.6363 | 89.3 | 9.2147 |
| 77.4 | 5.3463 | 81.4 | 6.4111 | 85.4 | 7.6711 | 89.4 | 9.2603 |
| 77.5 | 5.3712 | 81.5 | 6.4399 | 85.5 | 7.7060 | 89.5 | 9.3063 |
| 77.6 | 5.3961 | 81.6 | 6.4687 | 85.6 | 7.7411 | 89.6 | 9.3527 |
| 77.7 | 5.4212 | 81.7 | 6.4977 | 85.7 | 7.7764 | 89.7 | 9.3996 |
| 77.8 | 5.4463 | 81.8 | 6.5268 | 85.8 | 7.8120 | 89.8 | 9.4468 |
| 77.9 | 5.4715 | 81.9 | 6.5561 | 85.9 | 7.8477 | 89.9 | 9.4944 |
| 78.0 | 5.4967 | 82.0 | 6.5854 | 86.0 | 7.8837 | 90.0 | 9.5424 |
| 78.1 | 5.5221 | 82.1 | 6.6149 | 86.1 | 7.9199 | 90.1 | 9.5909 |
| 78.2 | 5.5475 | 82.2 | 6.6445 | 86.2 | 7.9563 | 90.2 | 9.6398 |
| 78.3 | 5.5730 | 82.3 | 6.6743 | 86.3 | 7.9929 | 90.3 | 9.6892 |
| 78.4 | 5.5986 | 82.4 | 6.7041 | 86.4 | 8.0297 | 90.4 | 9.7390 |
| 78.5 | 5.6243 | 82.5 | 6.7342 | 86.5 | 8.0668 | 90.5 | 9.7892 |
| 78.6 | 5.6501 | 82.6 | 6.7643 | 86.6 | 8.1041 | 90.6 | 9.8400 |
| 78.7 | 5.6760 | 82.7 | 6.7946 | 86.7 | 8.1417 | 90.7 | 9.8912 |
| 78.8 | 5.7019 | 82.8 | 6.8250 | 86.8 | 8.1795 | 90.8 | 9.9430 |
| 78.9 | 5.7279 | 82.9 | 6.8556 | 86.9 | 8.2175 | 90.9 | 9.9952 |
| 79.0 | 5.7541 | 83.0 | 6.8863 | 87.0 | 8.2558 | 91.0 | 10.0480 |
| 79.1 | 5.7803 | 83.1 | 6.9171 | 87.1 | 8.2943 | 91.1 | 10.1013 |
| 79.2 | 5.8066 | 83.2 | 6.9481 | 87.2 | 8.3331 | 91.2 | 10.1551 |
| 79.3 | 5.8330 | 83.3 | 6.9793 | 87.3 | 8.3721 | 91.3 | 10.2095 |
| 79.4 | 5.8595 | 83.4 | 7.0106 | 87.4 | 8.4114 | 91.4 | 10.2645 |
| 79.5 | 5.8861 | 83.5 | 7.0420 | 87.5 | 8.4510 | 91.5 | 10.3200 |
| 79.6 | 5.9128 | 83.6 | 7.0736 | 87.6 | 8.4908 | 91.6 | 10.3762 |
| 79.7 | 5.9396 | 83.7 | 7.1054 | 87.7 | 8.5309 | 91.7 | 10.4329 |
| 79.8 | 5.9665 | 83.8 | 7.1373 | 87.8 | 8.5713 | 91.8 | 10.4903 |
| 79.9 | 5.9935 | 83.9 | 7.1694 | 87.9 | 8.6120 | 91.9 | 10.5483 |

(*Contd.*)

| $P$, % | $db$ | $P$, % | $db$ | $P$, % | $db$ | $P$, % | $db$ |
|---|---|---|---|---|---|---|---|
| 92.0 | 10.6070 | 94.0 | 11.9498 | 96.0 | 13.8021 | 98.0 | 16.9020 |
| 92.1 | 10.6663 | 94.1 | 12.0274 | 96.1 | 13.9166 | 98.1 | 17.1292 |
| 92.2 | 10.7264 | 94.2 | 12.1062 | 96.2 | 14.0339 | 98.2 | 17.3684 |
| 92.3 | 10.7871 | 94.3 | 12.1864 | 96.3 | 14.1542 | 98.3 | 17.6210 |
| 92.4 | 10.8486 | 94.4 | 12.2678 | 96.4 | 14.2777 | 98.4 | 17.8888 |
| 92.5 | 10.9108 | 94.5 | 12.3507 | 96.5 | 14.4046 | 98.5 | 18.1734 |
| 92.6 | 10.9738 | 94.6 | 12.4350 | 96.6 | 14.5350 | 98.6 | 18.4775 |
| 92.7 | 11.0376 | 94.7 | 12.5207 | 96.7 | 14.6691 | 98.7 | 18.8037 |
| 92.8 | 11.1022 | 94.8 | 12.6080 | 96.8 | 14.8073 | 98.8 | 19.1558 |
| 92.9 | 11.1676 | 94.9 | 12.6970 | 96.9 | 14.9496 | 98.9 | 19.5380 |
| 93.0 | 11.2338 | 95.0 | 12.7875 | 97.0 | 15.0965 | 99.0 | 19.9564 |
| 93.1 | 11.3010 | 95.1 | 12.8798 | 97.1 | 15.2482 | 99.1 | 20.4183 |
| 93.2 | 11.3691 | 95.2 | 12.9740 | 97.2 | 15.4051 | 99.2 | 20.9342 |
| 93.3 | 11.4381 | 95.3 | 13.0700 | 97.3 | 15.5675 | 99.3 | 21.5185 |
| 93.4 | 11.5080 | 95.4 | 13.1679 | 97.4 | 15.7359 | 99.4 | 22.1924 |
| 93.5 | 11.5790 | 95.5 | 13.2679 | 97.5 | 15.9106 | 99.5 | 22.9885 |
| 93.6 | 11.6510 | 95.6 | 13.3701 | 97.6 | 16.0924 | 99.6 | 23.9620 |
| 93.7 | 11.7240 | 95.7 | 13.4744 | 97.7 | 16.2817 | 99.7 | 25.2157 |
| 93.8 | 11.7981 | 95.8 | 13.5812 | 97.8 | 16.4792 | 99.8 | 26.9810 |
| 93.9 | 11.8734 | 95.9 | 13.6903 | 97.9 | 16.6856 | 99.9 | 29.9957 |
| | | | | | | 100.0 | $\infty$ |

# References


Belavendram, N., 1995, *Quality by Design: Taguchi Techniques for Industrial Experimentation*, Prentice-Hall International, UK.

Box, G.E.P. and Behnken, D.W., 1960, Some New Three Level Designs for the Study of Quantitative Variables, *Technometrics*, Vol. 2, pp. 455–476, USA.

Box, G.E.P. and Draper, N.R., 1987, *Empirical Model Building and Response Surfaces*, John Wiley & Sons, New York.

Chao-Ton Su and Lee-Ing Tong., 1997, Multi-response Robust Design by Principal Component Analysis, *Total Quality Management*, Vol. 8, pp. 409–416, UK.

Cochran, W.G. and Cox, G.M., 2000, *Experimental Designs*, 2nd ed., John Wiley & Sons, Inc. India.

Draper, N.R. and Lin, D.K.J., 1990, *Small Response Surface Designs, Technometrics*, Vol. 32, pp. 187–194, USA.

Fisher, R.A. and Yates, F., 1953, *Statistical Tables for Biological, Agricultural and Medical Research*, 4th ed., Oliver and Boyd, Edinburg.

Goldberg, D.E., 1989, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, MA, USA.

Hung-Chang and Yan-Kwang, 2002, Optimizing Multi-response Problem in the Taguchi Method by DEA-based Ranking Method, *Int. J. of Quality and Reliability Management*, Vol. 19, pp. 825–837, UK.

Jeff Wu, C.F. and Michael Hamada, 2002, *Experiments: Planning, Analysis and Parameter Design Optimization*, John Wiley & Sons, Inc., India.

Jeyapaul, R., 2005, Simultaneous Optimization of Multi-response Taguchi Experiments, Unpublished Ph.D. Thesis, Anna University, Chennai, India.

Jeyapaul, R., Shahabudeen, P. and Krishnaiah, K., 2004, Quality Management Research by Considering Multi-response Problems in Taguchi Method—a Review. *Int. J. of Advanced Manufacturing Technology*, Vol. 26, pp. 1331–1337, UK.

__________, 2006, Simultaneous Optimization of Multi-response Problems in the Taguchi Method Using Genetic Algorithm, *Int. J. of Advanced Manufacturing Technology*, Vol. 30, pp. 870–878, UK.

Krishnaiah, K., 1987, Study and Modeling of Manual Lifting Capacity, Unpublished Ph.D Thesis, Anna University, Chennai, India.

Lin, C.L., Lin, J.L. and Ko, T.C., 2002, Optimization of EDM Process Based on the Orthogonal Array with Fuzzy Logic and Grey Relational Analysis Method, *Int. J. of Advanced Manufacturing Technology*, Vol. 19, pp. 271–277, UK.

Lin, J.L. and Tarng Y.S., 1998, Optimization of the Multi-response Process by the Taguchi Method with Grey Relational Analysis, *Journal of Grey Systems*, Vol. 10, pp. 355–370, China.

Mann, P.S., 2004, *Introductory Statistics*, 5th ed., John Wiley & Sons, India.

Mayers, R.H. and Montgomery, D.C., 1995, *Response Surface Methodology*, John Wiley & Sons, New York.

Montgomery, D.C., 2003, *Design and Analysis of Experiments*, 5th ed., John Wiley & Sons, India.

Panneerselvem, R., 2004, *Research Methodology*, Prentice-Hall of India, New Delhi, India.

Paul, G. Mathews, 2005, *Design of Experiments with MINITAB*, Pearson Education, India.

Phadke, M.S., 2008, *Quality Engineering Using Robust Design*, Dorling Kindersley, India.

Reddy, P.B.S., Nishina, K. and Babu, A.S., 1998, Taguchi's Methodology for Multi-response Optimization: A case study in the Indian plastic industry, *Int. J. of Quality and Reliability Management*, Vol. 16, pp. 646–668, UK.

Ross, P.J., 2005, *Taguchi Techniques for Quality Engineering*, 2nd ed., Tata McGraw-Hill, New Delhi.

Yates, F., 1937, The design and analysis of factorial experiments, *Imp. Bur. Soil Sci. Tech. Comm.* 35, UK.

Yenlay, O., 2001, A comparison of the performance between a Genetic Algorithm and the Taguchi method over artificial problems, *Turkish Journal of Engineering Environment Science*, Vol. 25, pp. 561–568, Ankara, Turkey.

# Index

# APPLIED DESIGN OF EXPERIMENTS AND TAGUCHI METHODS

**K. Krishnaiah • P. Shahabudeen**

Design of experiments (DOE) is an off-line quality assurance technique used to achieve best performance of products and processes. This book covers the basic ideas, terminology, and the application of techniques necessary to conduct a study using DOE.

The text is divided into two parts—Part I (Design of Experiments) and Part II (Taguchi Methods). Part I (Chapters 1–8) begins with a discussion on the basics of statistics and fundamentals of experimental designs, and then, it moves on to describe randomized design, Latin square design, Graeco-Latin square design. In addition, it also deals with a statistical model for two-factor and three-factor experiments and analyses $2^k$ factorial, $2^{k-m}$ fractional factorial design and methodology of surface design. Part II (Chapters 9–16) discusses Taguchi quality loss function, orthogonal design, and objective functions in robust design. Besides, the book explains the application of orthogonal arrays, data analysis using response graph method/analysis of variance, methods for multi-level factor designs, factor analysis and genetic algorithm.

This book is intended as a text for the undergraduate students of Industrial Engineering and postgraduate students of Mechatronics Engineering, Mechanical Engineering, and Statistics. In addition, the book would also be extremely useful for both academicians and practitioners.

## KEY FEATURES

- Includes **six** case studies of DOE in the context of different industry sectors.
- Provides essential DOE techniques for process improvement.
- Introduces simple graphical methods for reducing time taken to design and develop products.

## THE AUTHORS

**K. KRISHNAIAH**, PhD, is former Professor and Head, Department of Industrial Engineering, Anna University, Chennai. He received his BE (Mechanical) and ME (Industrial) from S.V. University, Chennai. Dr. Krishnaiah has twenty-eight years of teaching experience and has published several papers in national and international journals. His areas of interest include operations and production, design and analysis of experiments, and statistical quality control.

**P. SHAHABUDEEN**, PhD, is Professor and Head, Department of Industrial Engineering, Anna University, Chennai. He received his BE (Mechanical) from Madurai University and ME (Industrial) from Anna University, Chennai. Dr. Shahabudeen has published/presented more than thirty papers in international journals/conferences. His areas of interests include discrete system simulation, meta heuristics, object-oriented programming and operations research.




ISBN:978-81-203-4527-0

9 788120 345270

₹ 395.00

www.phindia.com