SECTION 47

DESIGN AND ANALYSIS OF EXPERIMENTS[1]

J. Stuart Hunter

with Mary G. Natrella, E. Harvey Barnett, William G. Hunter, and Truman L. Koehler

[1]This section borrows extensively from the Third and Fourth Editions, in which Mary Natrella helped prepare the section on design and analysis of experiments; E. Harvey Barnett helped prepare the section on evolutionary operation; and William G. Hunter and Truman L. Koehler helped prepare the section on response surface methodology. The present author, J. Stuart Hunter, gratefully acknowledges this earlier work and takes full responsibility for all changes in organization and emphasis.

INTRODUCTION

Experiments, statistically designed or not, are a component of the *learning process.* We experiment to learn. Learning through experimentation is a complex mechanism combining one's hopes, needs, knowledge, and resources. How well one succeeds will be a function of adherence to the *scientific method,* the most rapid means for speeding the learning process.

The scientific method is an iterative process. Ideas based upon one's current state of knowledge lead to experiments designed to answer questions. The experiments in turn lead to data that serve to confirm and modify these initial ideas. An iterative loop is established between ideas (*hypotheses*), the design of the experiment, the production of data, and the subsequent analysis (*inference*) leading on to new ideas, new experiments, new data, and newer inferences. This iterative process continues until an adequate state of knowledge is acquired and confirmed. The entire process is always slowed by "noise," the results of "errors" of measurement and experimental "variability." We "see through the glass darkly" and thus data analyses must be conditioned by statements of *uncertainty,* by considerations of probability. Statistics is the science, the provider of the language and logic, that combines the roles of hypotheses, data, inference, and uncertainty within the scientific method.

Two famous statisticians, G. E. P. Box and W. Edwards Deming, encapsulate the fundamentals of the scientific method into four segments: Box's "conjecture, design, experiment, analysis" and Deming's "plan, do, check, act." Both statements illustrate the role of statistics as an integral part of the scientific method; see Deming (1982), Box (1976), Ishikawa (1985).

In applying the scientific method, a worker's intelligence, experience and resources dominate the speed and success of any problem-solving program. The art of statistics comes into formal play in two places: in helping design the experiments to produce data heavily laden with information and in the data analysis wherein the statistician's responsibility is to uncover all the information of value. Simply stated: The application of the statistical design and analysis of experiments accelerates the learning process.

An experiment has been defined as a "considered course of action aimed at answering one or more carefully framed questions." Framing the question and planning for action are best accomplished through a team effort. An isolated experimenter working alone has become an increasingly rare phenomenon: the resources of information germane to any problem-solving and learning activity are too vast and varied to be left to a single individual.

Computing and Analysis. Almost 10 years have elapsed since the publication of the fourth edition of *Juran's Quality Control Handbook* (1988). This brief period has witnessed a spectacular growth in the use of the personal computer. Today the personal computer and associated software not only reduce the burdens of computations but can directly assist in the selection of an experimental

design. Computer-constructed graphical displays of data quickly assist in data analysis. The users manual accompanying many software programs is often equivalent to a reference book on the construction and analysis of statistical designs, offering help in the pathways of analysis and providing illustrative examples. Many statistical software programs go far beyond the subject of statistical design of experiments and include elements of industrial quality control, linear and nonlinear regression, reliability and nonparametric studies, and multivariate and time series analyses. The reader is encouraged to secure the use of a personal computer and software program capable of performing the analyses associated with the statistical design of experiments.

The challenge, of course, is to be wise in the use of the powerful tools of statistics and modern computation. This Handbook well serves as an introduction to the application of good statistical practices devoted to the pursuit of quality. As a further aid the Statistics Division of the American Society for Quality has produced a series of *How To* booklets (see ASQ in the references) designed to assist beginners in the applications of statistics. The best advice for those just starting out in the statistical design of experiments is KISS: Keep it simple statistician! One should first experience the design and analysis of a small program with limited amounts of data, perhaps a replicated balanced block to compare a few treatments, or a simple factorial to study the simultaneous influences of factors upon a response. Further, the beginner must recognize that there has never been a signal in the absence of "noise" and then learn to measure the noise: to estimate the variance of the observations. Beginners often confuse noise (errors) due to measurement and instrumentation with the more serious source of noise (error) caused by the failure of repeated experiments to provide identical results. Experimental error encompasses measurement error and is far more serious. A beginner will also quickly learn the importance of always plotting the data (Chambers et al. 1993; Cleveland 1993).

The methods of analysis and computation given here, and those found in most textbooks, are intended for hand or desk calculation. The modern personal computer and associated software dramatically change this environment, with some computer software programs providing levels of assistance that resemble artificial intelligence; the experimenter is asked questions and is guided through almost every stage of experimental design and analysis. Two general cautions are in order: (1) the hand-desk computing methods given here should not be literally translated into a computer program—there are alternative powerful computer-based methods of data reduction and analysis; and (2) the user of packaged statistical programs should be as critical a consumer of this as of everything else.

The journal *Applied Statistics* (The Royal Statistical Society, London) has in each issue a *Statistical Software Review* section providing careful and critical commentary on the utility of various software programs along with, usually, the reply of the producer of the software. The quarterly journal *The American Statistician,* of the American Statistical Association, has a section titled *Statistical Computing Software Reviews.* The reviews in both these journals can be of great value in judging the value of software. *Applied Statistics* also provides an additional section on *Statistical Algorithms* for those requiring special programs not commonly found in ordinary software. More general information on software for statistical purposes is provided in Section 44, Basic Statistical Methods. See also Wadsworth (1990).

Basic Definitions. Several fundamental terms are widely used throughout this section. They may be defined as follows:

Factor. A "factor" is one of the controlled or uncontrolled variables whose influence upon a response is being studied in the experiment. A factor may be quantitative, e.g., temperature in degrees, time in seconds. A factor may also be qualitative, e.g., different machines, different operators, switch on or off.

Level (Version). The "levels" ("versions") of a factor are the values of the factor being examined in the experiment. For quantitative factors, each chosen value becomes a level, e.g., if the experiment is to be conducted at four different temperatures, then the factor *temperature* has four *levels*. In the case of qualitative factors, *switch on or off* becomes two levels (versions) for the switch factor; if six machines are run by three operators, the factor *machine* has six levels (versions) while the factor *operator* has three levels (versions).

Treatment. A "treatment" is a single level (version) assigned to a single factor during an experimental run, e.g., temperature at 800 degrees. A "treatment combination" is the set of levels for all factors in a given experimental run. For example, an experimental run using an 800-degree temperature, machine 3, operator A, and switch off would constitute one treatment combination.

Experimental Units. The "experimental units" consist of the objects, materials, or units to which treatments are being applied. They may be biological entities, natural materials, fabricated products, etc.

Experimental Environment. The "experimental environment" comprises the surrounding conditions that may influence the results of the experiment in known or unknown ways.

Block. A factor in an experimental program that has influence as a source of variability is called a "block." The word is derived from early agricultural usage, in which blocks of land were the sources of variability. A block is a portion of the experimental material or of the experimental environment that is likely to be more homogeneous within itself than between different portions. For example, specimens from a single batch of material are likely to be more uniform than specimens from different batches. A group of specimens from such a single batch would be regarded as a block. Observations taken within a day are likely to be more homogeneous (to have smaller variance) than observations taken across days. *Days* then becomes a block factor.

Experimental Design. The formal plan for conducting the experiment is called the "experimental design" (also the "experimental pattern"). It includes the choice of the responses, factors, levels, blocks, and treatments and the use of certain tools called planned grouping, randomization, and replication.

Models. Experimental designs are created to help explain the association between a response variable η (η=eta; sometimes the symbol μ=mu is used) and other factors $x = x_1, x_2, x_3, \ldots, x_k$ (sometimes called "variables") thought to influence η. We say, "η (eta) is a function of x," that is, $\eta = f(x)$. Of course, *observing* the true response η entrains "noise" or "error" and produces observations $y = \eta + \epsilon$. The "expected value" of y is η. We are now faced with a "two model" problem: every observation y requires a model for η and another for ϵ.

The Model for the Error ϵ**.** The model for the ϵ usually assumed is that the errors are independent, normally distributed with an expected value of zero and a constant variance σ^2; that is, the errors are white Gaussian noise. One's ability to discern associations between a response of η and factors x and to make inferences about future performance are profoundly influenced by the ϵ, the "errors" (noise). The errors (noise) attending a series of experiments have two primary components, those due to measurement and those attending the repetition of the experiment. Considerations of *measurement* error; the precision, traceability, specificity, calibration, and cost of the measurements; and variability due to sampling are always important. However, measurement and sampling errors alone will underestimate the contribution of errors associated with running and then repeating an experiment. *Experimental* error, the failure of agreement between two or more separate experiments run under the same conditions, is of crucial importance. Experimental error includes measurement and sampling errors as components. Special experimental designs (the hierarchical designs, see below) can provide experimenters with the ability to measure separately all error components.

The assumption that the errors ϵ are normally distributed with zero expectation and constant variance is important. Transformations of experimental data are often required to accomplish these attributes. However, the assumption of independent errors is especially crucial whenever probability statements (hypothesis tests or confidence intervals) are made. To guarantee independence, acts of randomization should be part of every experimental design protocol. Randomization also provides support for the constant variance assumption.

The Model for the Response η**.** Implicit in every experiment is a *response model* $\eta = f(x)$ descriptive of how η changes as the factors x are changed. The specifics of this model are often not clear in

the mind of the experimenter; one object of the experiments is to reduce this ambiguity. Most experimental designs are constructed to provide sufficient data to estimate the parameters in a very general model, thus allowing the data to identify appropriate subsets of models for the experimenter's appraisal.

***The Model for the Observations* y.** The *observation model* is $y = \eta + \epsilon$. Thus the experimenter must separate motion among observations into two parts: one part assignable to η and ultimately to the changes in the factors x, and a second part assignable to ϵ, the noise. This separation of the roles of η and ϵ is formally recognized in an analysis of variance (ANOVA) table, an easy computation provided experiments are properly planned.

Tools for Good Experimentation. Good experimentation is an art and depends heavily upon the prior knowledge and abilities of the experimenter. Some tools of importance in the statistical planning and analysis of the design of experiments follow.

Blocking (Planned Grouping). Beyond the factors $x_1, x_2, \ldots, x_k$ selected for study, there are often other "background" variables that may also influence the outcome of the experimental program, variables such as raw material batches, operators, machines, or days. The influences of these variables upon the response are not under the control of the experimenter. These variables are commonly called "blocks," a legacy of the day when different blocks of land were used in agricultural experimentation. When an experimenter is aware of blocking variables it is often possible to plan experimental programs to reduce their influence. In designing experiments, wide use is made of the reduced variability occurring *within* blocks to accentuate the influences of the studied factors. Designs that make use of this uniformity within blocks are called "blocked" designs and the process is called "planned grouping."

Randomization. The sequence of experiments and/or the assignment of specimens to various treatment combinations in a purely chance manner is called "randomization." Such assignment increases the likelihood that the effect of uncontrolled variables will balance out. It also improves the validity of estimates of experimental error variance and makes possible the application of statistical tests of significance and the construction of confidence intervals. Whenever possible, randomization is part of the experimental program.

Replication. "Replication" is the repetition, the rerunning, of an experiment or measurement in order to increase precision or to provide the means for measuring precision. A single replicate consists of a single observation or experimental run. Replication provides an opportunity for the effects of uncontrolled factors or factors unknown to the experimenter to balance out and thus, through randomization, acts as a bias-decreasing tool. Replication also helps to detect gross errors in the measurements. In replications of groups of experiments, different randomizations should apply to each group.

Rerun experiments are commonly called "replicates." However, a sequence of observations made under a single set of experimental conditions, under a single replicate, are simply called "repeated observations."

Reproducibility and Repeatability. In manufacturing, reproducibility measures the variability between items manufactured on different days or on different machines. Repeatability measures sources of variability that are more local or immediate, assignable to item measurements or to the variability occurring between adjacent items manufactured in sequence.

Requisites and Tools. Table 47.1 lists some of the requisites for sound experimentation and shows the way in which these tools contribute to meeting these objectives. A checklist that can be helpful in all phases of an experiment is given in Table 47.2. Good references are Bicking (1954), Hahn (1977, 1984), Bisgaard (1992), Bishop et al. (1982), and Hoadley and Kettenring (1990), and Coleman and Montgomery (1993).

TABLE 47.1 Some Requisites and Tools for Sound Experimentation

Requisites	Tools
1. The experiment should have carefully defined objectives. See Table 26.2.	1. The definition of objectives requires all the specialized subject matter knowledge of the experimenter, and involves such things as: *a.* Choice of factors, including their range *b.* Choice of experimental materials, procedures, and equipment *c.* Choice of the metric for the factors (e.g., temperature or log temperature) and method of measurement
2. As far as possible, effects of factors should not be obscured by other variables.	2. The use of an appropriate experimental pattern helps to free the comparisons of interest from the effects of uncontrolled variables and simplifies the analysis of results.
3. As far as possible, the experiment should be free from bias, conscious or unconscious.	3. Some variables may be taken into account by planned grouping. Use randomization. Replication helps randomization to do a better job.
4. Experiments should provide a measure of experimental error variance (precision).	4. Replication provides the measure of variance and randomization ensures its validity.
5. Precision of experiment should be sufficient to meet objectives set forth in requisite 1.	5. Greater precision may be achieved by: refinements of measurement and experimental technique, experimental pattern (including planned grouping), replication.

CLASSIFICATION OF EXPERIMENTAL DESIGNS

Statisticians by themselves do not design experiments, but they have developed a number of structured schedules called "experimental designs," which they recommend for the taking of measurements. These designs have certain rational relationships to the purposes, needs, and physical limitations of experiments. Designs also offer certain advantages in economy of experimentation and provide straightforward estimates of experimental effects and valid estimates of variance. There are a number of ways in which experiment designs might be classified, for example, the following:

1. By the number of experimental factors to be investigated (e.g., single-factor versus multifactor designs)
2. By the structure of the experimental design (e.g., blocked, factorial, nested, or response-surface design)
3. By the kind of information the experiment is primarily intended to provide (e.g., estimates of effects, estimates of variance, or empirical mappings)

Some of the common statistical experimental designs are summarized in Table 47.3. Basic features of the designs are summarized in terms of these criteria of classification, and the details of design and analysis are given under the topics that follow. The analysis for observed responses is always based on a statistical model unique to the specific design.

TABLE 47.2 Checklist for Planning Test Programs

A. Obtain a clear statement of the problem.
   1. Identify the problem area in quantitative terms.
   2. Identify the response(s) to be measured, the factors that may be varied, the factors to be held constant, and the factors that cannot be controlled.
   3. Identify the ranges or limitations of the measurements and of the experimental factors.

B. Collect available background information.
   1. Investigate all available sources of information.
   2. Tabulate data pertinent to planning the experimental program.
   3. Be quantitative.

C. Design the experimental program.
   1. Hold a conference of all parties concerned.
      a. State the propositions to be explored.
      b. Agree on magnitude of differences in the response considered worthwhile.
      c. Outline possible alternative outcomes.
      d. Choose the factors to be studied.
      e. Determine practical range of factors and specify levels.
      f. Choose the measurements and methods of measurement.
      g. Consider the effect of sampling variability and of precision of the measurement methods.
      h. Consider possible interrelationships (interactions) of the factors.
      j. Determine influences of time, cost, materials, manpower, instrumentation, and other facilities and of extraneous conditions such as weather.
      k. Consider personnel and human relations requirements of the program.
   2. Design the experimental program in preliminary form.
      a. Prepare a systematic and inclusive schedule, which includes the randomization pattern.
      b. Provide for stepwise performance or adaptation of schedule if necessary.
      c. Eliminate effect of variables not under study by controlling, balancing, or randomization.
      d. Minimize the number of experimental runs consistent with objectives.
      e. Choose the method of statistical analysis.
      f. Arrange for orderly accumulation of data.
   3. Review the experimental design program with all concerned.
      a. Adjust the program as required.
      b. Spell out the steps to be followed in unmistakable terms.

D. Plan and carry out the experimental work.
   1. Develop methods, materials, and equipment.
   2. Carry out the experimental design in some random order.
   3. Record ancillary data.
   4. Record any modifications of the experimental design.
   5. Take precautions in the collection and recording of data, especially data from extra experiments and missing experiments.
   6. Record progress of the program by date, run number, and other ancillary data.

E. Analyze the data.
   1. Review the data with attention to recording errors, omissions, etc.
   2. Use graphics: plot the data, plot averages, plot simple graphs.
   3. Apply appropriate statistical techniques.

TABLE 47.2 Checklist for Planning Test Programs (*Continued*)

F. Interpret the results.
   1. Consider all the observed data.
   2. Confine initial conclusions to strict deductions from the experimental evidence at hand.
   3. Elucidate the analysis in both graphical and numerical terms.
   4. State results in terms of verifiable probabilities.
   5. Arrive at conclusions as to the technical meaning of results as well as their statistical significance.
   6. Point out implications of the findings for application and for further work.
   7. Account for any limitations imposed by the data or by the methods of analysis used.

G. Prepare the report.
   1. Describe work clearly, giving background, pertinence of problems, meaning of results.
   2. Use tabular and graphic methods of presenting data, and consider their possible future use.
   3. Supply sufficient information to permit readers to verify results and to draw their own conclusions.
   4. Limit conclusions to objective summary of evidence.

Split Plot. When certain (major) factors are difficult to change, other (minor) factors are run in an experimental design within each of the settings of the major factors. The Taguchi inner and outer array designs are of this genre. Many split-plot designs confound minor factor interactions with major factor main effects.

Major factors are not blocks. Major factors are studied for their main effects and interactions; blocks identify random variables. Further, blocks are assumed to have no interactions with the factors under study. When both the major and minor factors are random variables, the designs are identified as "nested."

COMPLETELY RANDOMIZED DESIGN: ONE FACTOR, k LEVELS

The completely randomized design is appropriate when a total of N experimental units are available for the experiment and there are k treatments (or levels of the factor) to be investigated. Of the total number N, it is usual to assign *randomly* an equal number of trials n to each of the k treatments.

Example. A study was made to investigate the effect of three different conditioning methods on the breaking strength T (in pounds per square inch) of cement briquettes. Fifteen briquettes were available from one batch and were assigned at random to the three methods. The results are summarized in Table 47.4. The purpose of the experiment was to investigate the effect of conditioning methods on breaking strength, and the analysis was designed to answer the question: Do the mean breaking strengths differ for the different conditioning methods?

This is an example of a randomized one-factor experiment. Only one experimental factor (method of conditioning) is under study. There are three methods; i.e., the number of treatments k equals 3. The number of units n assigned at random to each treatment is 5. The total number of experimental units N is 15.

Analysis. Almost everyone today possesses a computer and the software capable of performing the arithmetic associated with the analysis of most experimental designs. Once the data for the fully

TABLE 47.3 Classification of Designs

Design	Type of application	Structure	Information sought
Completely randomized	Appropriate when only one experimental factor is being investigated.	Basic: One factor is investigated by allocating experimental units at random to treatments (levels of the factor). Blocking: none.	1. Estimate and compare treatment effects. 2. Estimate variance.
Factorial	Appropriate when several factors are to be investigated at two or more levels and interaction of factors may be important.	Basic: Several factors are investigated at several levels by running all combinations of factors and levels. Blocking: none.	1. Estimate and compare effects of several factors. 2. Estimate possible interaction effects. 3. Estimate variance.
Blocked factorial	Appropriate when number of runs required for factorial is too large to be carried out under homogeneous conditions.	Basic: Full set of combinations of factors and levels is divided into subsets so that some high-order interactions are equated to blocks. Each subset constitutes a block. All blocks are run. Blocking: Blocks are usually units in space or time. Estimates of certain interactions are sacrificed to provide blocking.	1. Same as factorial except that certain high-order interactions cannot be estimated.

TABLE 47.3 Classification of Designs *(Continued)*

Design	Type of application	Structure	Information sought
Fractional factorial	Appropriate when there are many factors and levels and it is impractical to run all combinations.	Basic: Several factors are investigated at several levels but only a subset of the full factorial is run. Blocking: Sometimes possible.	1. Estimate and compare effects of several factors. 2. Estimate certain interaction effects (some may not be estimable). 3. Certain small fractional factorial designs may not provide sufficient information for estimating the variance.
Randomized block	Appropriate when one factor is being investigated and experimental material or environment can be divided into blocks or homogeneous groups.	Basic: Each treatment or level of factor is run in each block. Blocking: Usually with respect to only one variable.	1. Estimate and compare effects of treatments free of block effects. 2. Estimate block effects. 3. Estimate variance.
Balanced incomplete block	Appropriate when all the treatments cannot be accommodated in a block.	Basic: Prescribed assignments of treatments to blocks are made. Every pair of treatments will appear at least once in the experimental design, but each block will contain only a subset of pairs.	1. Same as randomized block design. All treatment effects are estimated with equal precision. Treatment averages must be adjusted for blocks.
Partially balanced incomplete block	Appropriate if a balanced incomplete block requires a larger number of blocks than is practical.	Basic: Prescribed assignments of treatments to blocks are made.	1. Same as randomized block design but all treatments are not estimated with equal precision.

TABLE 47.3 Classification of Designs *(Continued)*

Design	Type of application	Structure	Information sought
Latin square	Appropriate when one primary factor is under investigation and results may be affected by two other experimental variables or by two sources of nonhomogeneity. It is assumed that no interactions exist.	Basic: Two cross groupings of the experimental units are made corresponding to the columns and rows of a square. Each treatment occurs once in every row and once in every column. Number of treatments must equal number of rows and number of columns Blocking: With respect to two other variables in a two-way layout.	1. Estimate and compare treatment effects, free of effects of the two blocked variables. 2. Estimate and compare effects of the two blocked variables. 3. Estimate variance.
Youden square	Same as Latin square but number of rows, columns, and treatments need not be the same.	Basic: Each treatment occurs once in every row. Number of treatments must equal number of columns. Blocking: With respect to other variables in a two-way layout.	1. Same as Latin square.
Nested	Appropriate when objective is to study relative variability instead of mean effect of sources of variation (e.g., variance of tests on the same sample and variance of different samples).	Basic: Factors are strata in some hierarchical structure; units are tested from each stratum.	1. Relative variation in various strata, components of variance.

TABLE 47.3 Classification of Designs *(Continued)*

Design	Type of application	Structure	Information sought
Response surface	Objective is to provide empirical maps (contour diagrams) illustrative of how factors under the experimenter's control influence the response.	Factor settings are viewed as defining points in the factor space (may be multidimensional) at which the response will be recorded	Maps illustrating the nature of the response surface
Mixture designs	Same as factorial designs.	Many unique arrays. Factor settings are constrained. Factor levels are often percentages that must sum to 100%. Other factor level constraints are possible.	Same as factorial

TABLE 47.4 Breaking Strength T of Cement Briquettes, lb/in^2

	Method 1	Method 2	Method 3
	553	553	492
	550	599	530
	568	579	528
	541	545	510
	537	540	571
Total T	2749	2816	2631
n	5	5	5
Average $\overline{y}$	549.8	563.2	526.2
Estimate of variance s^2	145.7	626.2	864.2
Degrees of freedom	4	4	4

randomized design displayed in Table 47.4 have been placed into a computer with an appropriate statistical design of experiments software program, then graphics similar to Figure 47.1 and an analysis of variance table such as that displayed in Table 47.5 become available. Nevertheless, we include here the details of the computations for the reader. The ability to perform one's own hands-on calculations is often of great value in reducing the "black-box" approach to data analysis.

The analysis of these data begins with a plot of the three treatment averages as shown in Figure 47.1. The "reference distribution" for these averages will be explained shortly. The average responses

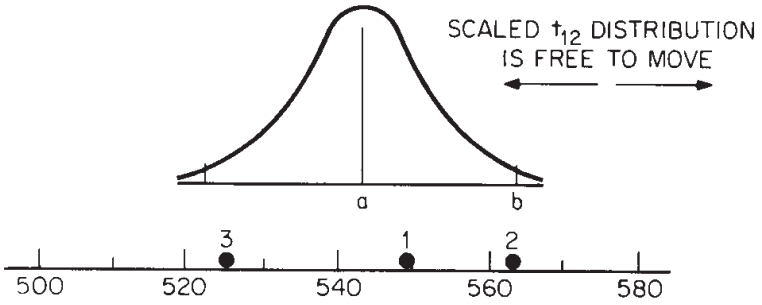


FIGURE 47.1 Plot of treatment averages and their reference *t*-distribution. (Distance from a to $b = t_{\nu,\,\alpha/2}\sqrt{s^2/n} = t_{12,\,0.025}\sqrt{s^2/n} = 2.179\ \sqrt{545.4/5} = 22.8$.)

TABLE 47.5 Analysis of Variance for a One-Factor Experiment

Source of variation	Sum of squares	Degrees of freedom	Mean square
Between treatments	SSB = 3,509.2	$(k-1) = 2$	MSB = SSB/$(k-1)$ MSB = 3509.2/2 = 1754.6
Within treatments	SSW = 6,544.4	$k(n-1) = 12$	MSW = SSW/$k(n-1)$ MSW = 6544.4/12 = 545.4
Total	TSS = 10,053.6	$(N-1) = 14$	

are obviously different. The key question is whether the observed differences are due solely to the inherent variability of the observations or caused by this variability plus real differences between the treatment means. (In this section on experimental design the word "mean" is used to connote the *expected value* of an average, that is, the value an average would take if an infinite number of observations were made.) The analysis of variance (ANOVA) is a basic statistical technique in the analysis of such data and is illustrated for the data in Table 47.5.

Analysis of Variance. Referring to Table 47.4, the total T is

$$T = 8196 \qquad N = 15 \qquad n_i = 5 \qquad i = 1, 2, 3$$

Calculate the following:

The *uncorrected* total sum of squares $\Sigma y^2 = 4{,}488{,}348$

$$C = \text{correction factor (a special constant)} = \frac{T^2}{N} = \frac{(8196)^2}{15} = 4{,}478{,}294.4$$

TSS = *corrected* total sum of squares

$$= \Sigma y^2 - C = 4{,}488{,}348 - 4{,}478{,}294.4 = 10{,}053.6$$

SSB = between-treatments sum of squares

$$= \sum_i \frac{T_i^2}{n} - C = \frac{22{,}409{,}018}{5} - C = 4{,}481{,}803.6 - 4{,}478{,}294.4 = 3509.2$$

SSW = within-treatments sum of squares = TSS − SSB = 10,053.6 − 3509.2 = 6544.4

(SSW is here obtained by subtraction. It may be obtained directly for the special case of the completely randomized design by calculating $\Sigma y^2 - [(\Sigma y)^2/n]$ for each treatment and summing for all treatments.)

The corrected total sum of squares TSS has $N - 1$ degrees of freedom (DF), the between-treatments sum of squares SSB has $(k - 1)$ degrees of freedom, and the within-treatments sum of squares SSW has $(N - k)$ degrees of freedom.

The mean-square column in the table is computed as follows:

$$\text{Mean square between treatments MSB} = \frac{\text{SSB}}{k - 1}$$

$$\text{Mean square within treatments MSW} = \frac{\text{SSW}}{N - k}$$

These calculated quantities are inserted in Table 47.5.

Calculate $F = \text{MSB/MSW} = 1754.6/545.5 = 3.22$. Choose α, the significance level of the test. If the calculated F exceeds $F_{1-\alpha}$ from Appendix II, Table K, for $(k - 1)$ and $k\,(n - 1)$ degrees of freedom, conclude that there are differences among treatment means. For example, $\alpha = 0.05$ level, $F_{0.95}$ for (2,12) degrees of freedom = 3.89; the calculated F does not exceed this value, and we do not have sufficient evidence to conclude that the mean breaking strength is different for the different conditioning methods. The differences between the treatment averages are thus assumed to be due to the error variance (the noise). Had the hypothesis that there are no treatment effects been rejected by the F test, then one could conclude that at last one of the mean breaking strengths differs from the others. The formal issue of comparing treatment means, the problem of "multiple comparisons," is discussed below.

In this example, the n_i are all equal. Designs that have an equal number of observations in each treatment are generally to be preferred. Such designs provide each treatment with an equal opportunity for comparison against all other treatments. On rare occasions, as when one of the treatments is a standard against which all other treatments are to be compared, more observations are placed in the standard treatment than in the alternatives. When the n_i are not all equal, use the following formula for the between-treatments sum of squares:

$$\text{SSB} = \frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \ldots + \frac{T_k^2}{n_k} - C$$

Here SSW = TSS − SSB and MSW = SSW/$(N - k)$ as before. The MSW (the overall estimate of variance) can also be obtained by estimating the variance within each treatment and pooling these estimates. The pooled estimate has $N - k$ degrees of freedom.

Graphical Analysis. An approximate graphical analysis of these data is provided by sketching an appropriate "reference" distribution for the average, as illustrated in Figure 47.1. When σ^2 (the population variance) or, equivalently, σ (the population standard deviation) is known, averages may be referred to a normal distribution with standard error $\sigma/\sqrt{n}$. Here σ^2 is unknown, and its estimate s^2 must be determined from the data. Averages must be referred to a Student's t-distribution, suitably scaled by $s/\sqrt{n}$. The number of degrees of freedom in t is equal to the number of degrees of freedom in s^2. The estimate of σ^2 is obtained from the analysis of variance table (s^2 = MSW) or by pooling the separate estimates of variance obtained from within each treatment classification. The pooled estimate of variance is given by the weighted average of the individual estimates, the weights being their degrees of freedom. Thus

$$s^2 = \sum_i \frac{(n_i - 1)s_i^2}{N - k}$$

$$= \frac{4(145.7) + 4(626.2) + 4(864.2)}{16} = 545.5$$

$$s = 23.34 \text{ with 12 degrees of freedom}$$

A sketch of the appropriately scaled t-distribution appears in Figure 47.1. The distance from the center of the t-distribution to each extremal point equals $t_{12,\,0.025}s/\sqrt{n} = 2.18(23.35)/\sqrt{5} = 22.8$, and 95 percent of the distribution falls within the range $2(22.8) = 45.6$. The distribution is easily sketched; there is no need for great precision in its shape save that it look reasonably bell-shaped. The distribution can be moved back and forth, and has been located in this instance so that it just fits over the three averages. Thus, based on this graphical evidence, the suggestion that all three averages could have come from the same parent distribution and hence could be estimates of the identical mean seems reasonable.

This graphical interpretation is confirmed by using the F test in the analysis of variance table; that is, the computed F ratio was not statistically significant. When averages do not fit reasonably under the distribution, the graphical analysis suggests that the differences between the treatment averages reflect real differences between the treatment means. Such graphical evidence could, of course, be verified by an F test. The scaled reference distribution can be of great value in interpreting treatment averages. Often, although it is technically possible to place the reference distribution over all the averages and simultaneously to obtain a nonsignificant F test, interesting differences between groups of the averages may become clear. The F test does not consider patterns among the averages, a factor that can be of great importance to the analyst (see Box, Hunter, and Hunter 1978). Alternative approximate analysis techniques, in particular those based upon the use of the range statistic in place of the estimated standard deviation, are available [see the papers by Kurtz et al. (1965) and by Sheesley (1980)].

GENERALIZED COMMENTS ON A COMPLETELY RANDOMIZED DESIGN

The completely randomized design is simple to organize and analyze and may be the best choice when the experimental material is homogeneous and when background conditions can be well controlled during the experiment.

The advantages of the design are

1. Complete flexibility in terms of number of treatments and number of units assigned to a treatment
2. Simple analysis
3. No difficulty with lost or missing data

In planning the experiment, n units are assigned at random to each of the k treatments. When the data have been taken, the results are set out as in Table 47.6.

TABLE 47.6 Completely Randomized Design

Observations within treatments*	Treatments				
	1	2	3	· · ·	k
1	y_{11}	y_{12}	y_{13}	· · ·	y_{1k}
2	y_{21}	y_{22}	y_{23}	· · ·	y_{2k}
3	y_{31}	y_{32}	y_{33}	· · ·	y_{3k}
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
n	y_{n1}	y_{n2}	y_{n3}	· · ·	y_{nk}

*The entire nk observations are recorded in random order *without* regard to the treatment classifications.

Displayed this way, the results of experiments are indistinguishable from a situation in which there has been no design and no allocation at all but in which several different samples have been tested from each of several different sources of material or several observations have been made under each of several different conditions. Whether the observations come from units randomly allocated to several different treatments or from units obtained from several different sources, the data table looks the same, and in fact the analysis will be essentially the same.

This simple one-factor design is called "completely randomized" to distinguish it from other experiment designs in which either randomization is constrained or the principle of "blocking," or planned grouping, has been made part of the structure.

One-Way Analysis of Variance—Models. The results of an experiment run according to a completely randomized design are summarized in a one-way table such as Table 47.6. The completely randomized design is called a "one-way" classification of data, whether or not the data came from a designed experiment. To discuss the associated analysis of variance, statisticians require "models" for the data. In the case of a one-way classification analysis of variance, the most appropriate model is determined by answering the question: Do the several groups (into which the data are classified) represent *unique* groups of interest to the experimenter? If they do, the model is called "Model I," the "Fixed Effects Model." If, on the other hand, the groups are considered to be a random sample from some population made up of many such groups, the model is called "Model II," the "Random Effects Model." For example, suppose that the data in Table 47.4 were not from a completely randomized design in which 15 briquettes were allocated at random to three unique conditioning treatments of interest to the experimenter. Suppose instead the column headings were "Batch 1," "Batch 2," "Batch 3," where the "batches" represented some convenient grouping of briquettes so that five briquettes were tested per batch. In the original experimental program, in which the three conditioning treatments of the designed experiment were the unique treatments of interest to the experimenter, the data may be represented by the Fixed Effects Model (Model I), whereas in the second program the three batches presumably were a random sample of batches, and hence these data are represented by the Random Effects Model (Model II).

For both Model I and Model II, the experimenter is trying to determine whether the three groups are different in mean value. For Model II the experimenter may also be interested in knowing about the "components of variance"; that is, the variance between samples from the same batch and the variance existing between batches. Knowledge of the variability of different samples within a batch and between different batches is helpful in planning how many samples to test in future experiments.

Data obtained from a designed experiment, as described for this completely randomized design, are usually considered to be represented by Model I, since presumably the experimenter includes the treatments of interest. If the data correspond to Model II, the analysis of variance table and F test are used with one extra step, which requires adding an extra column labeled "Expected Mean Square," to Table 47.5.

	Expected mean square
Between groups	$\sigma_w^2 + n\sigma_b^2$
Within groups	σ_w^2

For the data in Table 47.5, we have:

	Mean square	Expected mean square
Between groups	MSB=1754.6	$\sigma_w^2 + 5\sigma_b^2$
Within groups	MSW=545.4	σ_w^2

The quantity σ_b^2 is called the "*between* component of variance," and σ_w^2 is called the "*within* component of variance"; MSB is an estimate of the "Expected Between-Groups Mean Square"; and MSW is an estimate of the "Expected Within-Groups Mean Square." Estimates of s_b^2 and s_w^2 or σ_b^2 and σ_w^2 can be obtained as follows:

MSB is set equal to $s_w^2 + ns_b^2$ and MSW is set equal to s_w^2:

$$s_w^2 = 545.4$$

$$s_b^2 = \frac{1754.6 - 545.4}{5} = \frac{1209.2}{5} = 241.8$$

The total variance assignable to a single observation from one randomly chosen batch and a single briquette is estimated as

$$s^2 = s_b^2 + s_w^2$$

BLOCKED DESIGNS

The several levels, or versions, of a studied factor or group of studied factors are called "treatments," and the major objective of an experimenter is to study the influences of these different treatment levels upon some response. Often all the levels of the studied factors are repeated each day or with a different operator, machine, supply of raw materials, etc. Each complete replication of the set of treatments is called a "block." The experimenter should plan the treatments so as to prevent differences between the blocks from influencing the comparisons between the treatments. For example, if the blocks in the experiment are days, the first aim of the experiment is to evaluate the effects of the studied factors free of the effects of day differences. A secondary aim might actually be to measure the effects of the days to help in planning future experiments. In blocked designs it is generally assumed that blocking factors do not interact with studied factors. In the simplest block designs the data, when taken, can be summarized in a two-way table, as illustrated in Table 47.7. Note that this design is *not* a factorial design. In a factorial design all the factors (here rows and columns) are at predetermined levels, whereas in the design under consideration here, the blocking factor is not under the control of the experimenter. The blocking factor is, however, recognized as capable of influencing the response. The experimenter's objective is to remove from the influences of the studied factors any possible contributions to the response that are provided by the blocking factors. Blocking factors are commonly environmental phenomena outside the control of the experimenter.

The interest in the factor called blocks has several objectives. Some of these are:

TABLE 47.7 Schematic for a Simple Block Design

	Treatments			
	1	2	· · ·	k
Block 1	y_{11}	y_{12}	· · ·	y_{1k}
Block 2	y_{21}	y_{22}	· · ·	y_{2k}
·	·	·	·	·
·	·	·	·	·
·	·	·	·	·
Block b	y_{b1}	y_{b2}	· · ·	y_{bk}

1. The aim of the experimenter is to estimate effects of treatments free of block effects; numerical estimates of block effects are not particularly needed. For example, if blocks are days, day-to-day differences should be eliminated as sources of variability and are of no particular interest in themselves.
2. The primary aim is to estimate effects of the treatments (the studied factors) and secondarily to have estimates of block effects.
3. Sometimes the treatment effects and the block effects are of almost equal interest. In this case "block design" is analogous to a "two-factor experiment," but the experimenter must be sure that the studied and blocking factors do not interact before using a block design data analysis. If interaction between factors exists or is suspected, the design and analysis for a factorial experiment must then be used.

The simplest design with one-way blocking is the "randomized block design."

RANDOMIZED BLOCK DESIGN

In comparing a number of treatments, it is clearly desirable that all other conditions be kept as nearly constant as possible. Unfortunately, the required number of tests is often too large to be carried out under similar conditions. In such cases, the experimenter may be able to divide the experiment into blocks, or planned homogeneous groups. When each such group in the experiment contains exactly one observation on every treatment, the experimental plan is called a randomized block plan. The treatments are run *in a random order* within the blocks.

There are many situations in which a randomized block plan can be profitably utilized. For example, a comparison of several levels of some factor may take several days to complete. If we anticipate that the different days may also have an influence upon the response, then we might plan to observe all of the factor levels on each day. A day would then represent a block. In another situation, several operators may be conducting the tests, and differences between operators may be expected. The tests or observations made by a given operator can be considered to represent a block. The size of a block—that is, the number of tests contained within the block—may be restricted by physical considerations. In general, a randomized block plan is one in which each of the treatments appears exactly once in every block. The treatments are allocated to experimental units at random within a given block. The results of a randomized block experiment can be exhibited in a two-way table such as Table 47.8, in which we have $b = 4$ blocks and $k = 6$ treatments. Since each treatment occurs exactly once in every block, the treatment totals or averages are directly comparable without adjustment.

Example. The data in Table 47.8 represent the conversion gain of four resistors measured under six different conditions. The response, conversion gain, is defined as the ratio of available current-noise power to applied dc power expressed in decibel units and is a measure of the efficiency with which a resistor converts dc power to available current-noise power. Each test condition involves the same four resistors. The experimenter is interested in comparing differences between conditions (the studied factor) clear of possible influences due to the resistors (the blocking factor). A quick review of Table 47.8 indicates large differences between the resistors, i.e., between the block averages. The key question is whether, with this resistor variability eliminated, the experimenter can now detect real differences between the test conditions since the differences between the observed condition averages are small and may merely reflect experimental error.

Analysis. Some computer software programs may title the analysis of variance associated with the randomized block experiment as a "two-way" analysis of variance. In a randomized block experiment primary interest rests in the treatment averages and their standard errors. Block averages are always of interest, but since blocks cannot be controlled by the experimenter (they represent environ-

TABLE 47.8 Randomized Block Design Response: Conversion Gain of Resistors

Resistor (blocks)	Test set (treatments) 1	2	3	4	5	6	Row total	Row average
1	138.0	141.6	137.5	141.8	138.6	139.6	$B_1 = 837.1$	$b_1 = 139.52$
2	152.2	152.2	152.1	152.2	152.0	152.8	$B_2 = 913.5$	$b_2 = 152.25$
3	153.6	154.0	153.8	153.6	153.2	153.6	$B_3 = 921.8$	$b_3 = 153.63$
4	141.4	141.5	142.6	142.2	141.1	141.9	$B_4 = 850.7$	$b_4 = 141.78$
Column totals								
$T_1 = 585.2$	$T_2 = 589.3$	$T_3 = 586.0$	$T_4 = 589.8$	$T_5 = 584.9$	$T_6 = 587.9$		Grand total $G = 3523.1$	
Column averages								
$\bar{y}_1 = 146.30$	$\bar{y}_2 = 147.32$	$\bar{y}_3 = 146.50$	$\bar{y}_4 = 147.45$	$\bar{y}_5 = 146.22$	$\bar{y}_6 = 146.98$			

ments within which the treatments have been randomly run) their importance is commonly secondary to that of the treatments. Of course, comparisons between averages requires an estimate of the variance σ^2. The primary reason for the analysis of variance computations is to get an estimate of the variance (to quantify the noise) clear of all assignable causes. Plots of the averages are always required.

The analysis of a randomized block experiment depends on a number of assumptions. We assume that each of the observations is the sum of four components. If we let y_{ij} be the observation on the ith treatment in the jth block, then

$$y_{ij} = \eta + \phi_i + \beta_j + \epsilon_{ij}$$

The term η is the grand mean, ϕ_i is the effect of treatment i, β_j the effect of block j, and ϵ_{ij} the experimental error associated with the measurement y_{ij}. (The subscripts $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, b$.) The mean for the ith treatment equals $\eta + \phi_b$ and the mean for the jth block equals $\eta + \beta_j$. The terms ϕ_i and β_j represent, respectively, the unique contributions (effects) of treatments and blocks. The estimate of the mean η is given by $\bar{y}$, the grand average. Letting $\bar{y}_i$ equal the average for the ith treatment, the estimate of treatment effect ϕ_i is $\bar{y}_i - \bar{y}$. Similarly, $\bar{y}_j - \bar{y}$ estimates the block effect β_j.

In order to make interval estimates for or tests of hypotheses on the treatment or block contributions, we assume that the values of the experimental error ϵ_{ij} are independently and normally distributed with constant variance. If the experiment is randomized properly, failure of these assumptions will, in general, not cause serious difficulty.

A more serious difficulty occurs when count data are recorded. Count data are frequently Poisson-distributed, and hence the variance of the observations is linked directly to their mean. In such circumstances, it is best first to take the square roots of the count data and then to proceed with the estimation of effects and the analysis of variance.

Reference Distribution for Treatment Averages. The plot of the $k = 6$ treatment averages is displayed in Figure 47.2

To construct the appropriate reference distribution to judge these averages and to test hypotheses, an estimate of the experimental error variance σ^2 is required. Using the model, the associated analysis of variance table can now be constructed.

Randomized Block Analysis of Variance. The analysis of variance table for this randomized block experiment data with $k = 6$ treatments and $b = 4$ blocks is given in Table 47.9.

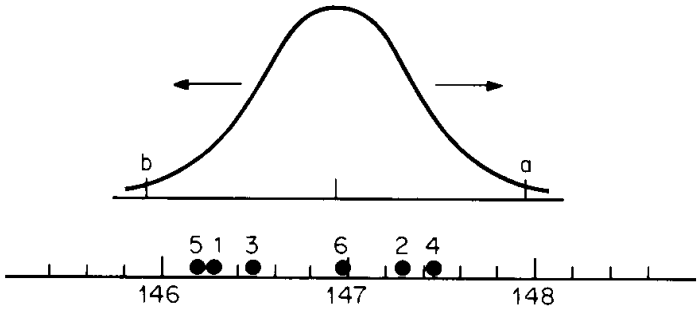


FIGURE 47.2 Plot of $k=6$ treatment averages and their reference t-distribution. (Distance from a to $b = 2ts/\sqrt{b} = 2.02$.) Numbers above scale are treatment numbers from Table 47.8.

TABLE 47.9 Analysis of Variance Table: Randomized Block Design

	Sum of squares (SSq)	Degrees of freedom (DF)		Mean square
Blocks: SSB	927.665	$(b-1)$	3	309.222
Treatments SST	5.598	$(k-1)$	5	1.120
Residual SSR	13.467	$(b-1)(k-1)$	15	$0.898 = s^2$
Total corrected SSq	946.730	$(bk-1)$	23	

The details of these computations are as follows:

Let $N = bk$ = total number of observations. Here $N = (4)(6) = 24$.

Let G = grand total of all the observations. $G = 3523.1$.

Let Σy_{ij}^2 = sum of the $y_{ij}^2 = 518{,}123.13$.

Total corrected SSq $= \Sigma y_{ij}^2 - G^2/N = 518{,}123.13 - (3523.1)^2/24$.

Let B_i = total for block i, $i = 1, 2, \ldots, b$. Here $b = 4$.

Blocks: SSB $= \Sigma_i B_i^2/k - G^2/N = 518{,}104.065 - G^2/N = 927.665$.

Let T_j = total for treatment j, $j = 1, 2, \ldots, k$. Here $k = 6$.

Treatments SST $= \Sigma_j T_j^2/b - G^2/N = 517{,}181.998 - G^2/N = 5.598$.

The residual SSR is obtained by subtraction.

The "mean squares" = (sum of squares)/(degrees of freedom).

An excellent explanation of these computations can be found in Box, Hunter, and Hunter (1978).

To test the hypothesis that all the treatment means are equal select an α risk level (commonly 0.05) and perform an $F_{\alpha,\nu 1,\nu 2}$ test. Thus

$$F_{\nu 1,\nu 2} = F_{5,15} = \frac{\text{SST}/(k-1)}{s^2} = \frac{1.120}{0.898} = 1.247$$

The critical $\alpha = 0.05$ value of $F_{0.05,5,15} = 2.90$. (Critical values of the F ratio are found in Table K, Appendix II.) The computed F is less than the critical F and we therefore declare there is insufficient evidence to lead to the rejection of the hypothesis that all treatment effects are zero. We may not actually believe this hypothesis, but we cannot reject it. A similar test that all block effects are zero gives:

$$\text{F}_{\nu 1,\nu 2} = \text{F}_{3,15} = \frac{\text{SSB}/(b-1)}{s^2} = \frac{309.222}{0.898} = 344.345$$

The computed F is far greater than the critical $F_{0.05,3,15} = 3.29$, and the hypothesis that there are zero block effects is rejected.

Figure 47.2 shows the plot of the six treatment averages and their associated "reference distribution." The reference distribution for averages is a normal distribution scaled by $\sqrt{\sigma^2/n}$ when σ^2 is known. When, as in this case, only $s^2 = 0.898$ with $\nu = 15$ degrees of freedom is known, the refer-

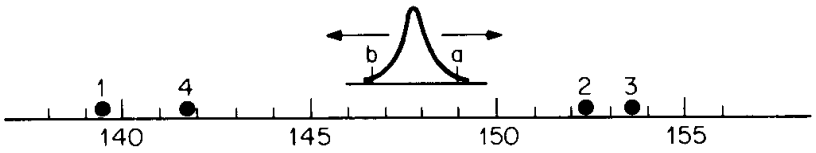


FIGURE 47.3 Plot of $b = 4$ block averages from Table 47.8 and their reference t-distribution. (Distance from a to $b = 2ts/\sqrt{k} = 1.65$.)

ence distribution for averages becomes a t-distribution scaled by $\sqrt{s^2/n}$. The distribution displayed in Figure 47.2 is the bell-shaped t curve, 95 percent of its area contained within the interval

$$2t_{\alpha,\nu}\sqrt{s^2/n} = 2(2.131)\sqrt{(0.898)/4} = 2.02$$

(The number of observations in a treatment average is $n = b = 4$.) The curve can be sketched in by hand; great precision is not required. One imagines this bell-shaped t-distribution to move back and forth horizontally. If two or more averages appear to be nested under a single location of the distribution, the inference is that differences between these averages may be due solely to the variance of the observations, a graphical test analogous to a nonsignificant F test. If some of the averages, individually or in clusters, seem to aggregate beyond the bell curve, this is taken as a signal that real differences exist, analogous to a significant F test.

Figure 47.3 displays the $b = 4$ block averages and their reference t-distribution. Once again, 95 percent of the distribution contained in the interval is given by

$$2t_{\alpha,\nu}\sqrt{s^2/n} = 2(2.131)\sqrt{(0.898)/6} = 1.65$$

(The number of observations in a block average is $n = k = 6$.) It is obvious from viewing Figure 47.3 that the reference distribution, wherever horizontally positioned, cannot reasonably account for the four averages. Block averages 1 and 4 are distinctly different from 2 and 3 and from one another. There is some indication, though slight, that averages 2 and 3 may be from a single distribution and hence merely reflections of the variability of averages about their expected mean value.

MULTIPLE COMPARISONS

An alternative graphical device for comparing treatment averages is the *analysis of means,* originated by Ott (1967). A plot similar to the Shewhart control chart is constructed. The control limits for the charts are easily obtained using special tables that adjust for the multiple comparisons that are possible. See Ott (1967), Schilling (1973), Nelson, L. (1983), Ott and Schilling (1990), Nelson, P. (1993), and, for good examples, Ramig (1983).

The graphical technique of a sliding reference distribution is subjective, the experimenter's eye and good judgment called into play. Many alternative formal techniques are available to demonstrate whether differences exist between the treatment means. Commonly, special intervals are computed such that averages appearing together within an interval may not be declared statistically significantly different. One approach is to compute the *least significance difference* (LSD) where

$$\text{LSD}_\alpha = t_{\nu,\alpha/2}\sqrt{\left(\frac{1}{n_i} + \frac{1}{n_j}\right)s^2}$$

If the absolute value of the difference between any two averages is greater than the LSD, the two treatment averages may be declared statistically significantly different. Of course, in comparing k averages there are $k(k-1)/2$ possible pairs, and many other comparisons (contrasts) are possible when k is modestly large. In LSD comparisons a fixed α risk (usually $\alpha = 0.05$) is maintained for each comparison regardless of the number of comparisons made. The overall α risk is considerably

increased when multiple tests are performed. For a fixed overall α risk for all possible *pairs* of treatments, one may wish to use the Tukey Studentized range statistic (Tukey 1949). Dunnett's method (Dunnett 1964; Bechofer and Tamhane 1983) is used for all $(k - 1)$ differences from a standard. The Scheffé test is used to make *all* possible comparisons (Scheffé 1953). Other multiple comparison methods are the Bonferroni interval (Dunn and Clark 1987); and the Duncan multiple range (Duncan, D. B., 1955). The Bonferroni and Duncan multiple range procedures can be found in most computer software programs. A good overall reference to the problems of multiple comparisons is Miller (1981). For multiple comparison of variances see Spurrier (1992).

BALANCED INCOMPLETE BLOCK DESIGNS

In an incomplete block design, all the treatments cannot be accommodated within a single block. To illustrate, consider a production manager who wishes to study the differences between the products supplied by six different suppliers. Unfortunately, personnel and equipment limit the number of suppliers that can be studied to three a day. The production manager is concerned that day-to-day differences might upset comparisons between suppliers and wishes to block the contributions of days, but the individual blocks are not large enough to encompass all six treatments. The appropriate experimental design to use is a "balanced incomplete block design," as illustrated in Table 47.10. The six suppliers (treatments) labeled A, B, C, D, E, and F are then studied in groups of three within each day (block). The blocks and the sequence of trials within the blocks are to be chosen in some random order.

Note that in the design displayed in Table 47.10 every letter supplier is tested the same number of times and every pair of letters appears within a block the same number of times. Another design appropriate to the case of six treatments constrained to be studied three at a time is the "combinatoric" design, that is, a design consisting of all combinations of six things taken three at a time. The combinatoric balanced incomplete block would have required 20 blocks; the design illustrated in Table 47.10 requires only 10.

One consequence of using an incomplete block design is that each treatment average must be adjusted for the blocks in which it appears and the differences between the *adjusted* treatment averages appraised. The computations are straightforward but go beyond what can be accommodated in this handbook. Interested readers are referred to Cochran and Cox (1957), Natrella (1963), and Box, Hunter, and Hunter (1978).

To enumerate the situations in which it is possible to construct a balanced incomplete block design, the quantities *r, b, t, k, L, E,* and *N* are defined as follows:

r = number of replications (the number of times a treatment appears)

b = number of blocks in the plan

TABLE 47.10 A Balanced Incomplete Block Design

Days	Before randomization	After randomization	Six treatments A B C D E F
1	ABE	DBC	t = 6 treatments
2	ABF	CDA	b = 10 blocks
3	ACD	FBD	k = 3 treatments/block
4	ACF	CEB	r = 5 replicates/treatment
5	ADE	DEF	L = 2, i.e., each
6	BCD	EBA	treatment pair
7	BCE	AFB	appears twice
8	BDF	DEA	
9	CEF	FCE	
10	DEF	CFA	

t = number of treatments

k = block size, i.e., the number of treatments that can appear in each block

L = number of blocks in which a given treatment pair appears: $L = r(k - 1)/(t - 1)$

E = a constant used in the analysis: $E = tL/rk$

N = total number of observations: $N = tr = bk$

Plans are indexed in Table 47.11 for $4 < t < 10$ and $r < 10$. For an extensive listing of the designs and many worked examples see Cochran and Cox (1957) and Natrella (1963).

General Comments on Block Designs. In the simplest type of block design, Randomized Blocks, each block is large enough to accommodate all the treatments one wishes to test. In Incomplete Block Designs, the block size is not large enough for all treatments to be tested in every

TABLE 47.11 Balanced Incomplete Block Plans ($4 < t < 10$ and $r < 10$)

t	k	r	b	L	E*	Plan†
4	2	3	6	1	2/3	1
	3	3	4	2	8/9	Comb‡
5	2	4	10	1	5/8	2
	3	6	10	3	5/6	Comb
	4	4	5	3	15/16	Comb
6	2	5	15	1	3/5	3
	3	5	10	2	4/5	4
	3	10	20	4	4/6	5
	4	10	15	6	9/10	6
	5	5	6	4	24/25	Comb
7	2	6	21	1	7/12	Comb
	3	3	7	1	7/9	7
	4	4	7	2	7/8	8
	6	6	7	5	35/36	Comb
8	2	7	28	1	4/7	9
	4	7	14	3	6/7	10
	7	7	8	6	48/49	Comb
9	2	8	36	1	9/16	Comb
	3	4	12	1	3/4	11
	4	8	18	3	27/32	12
	5	10	18	5	9/10	13
	6	8	12	15	15/16	14
	8	8	9	7	63/64	Comb
10	2	9	45	1	5/9	15
	3	9	30	2	20/27	16
	4	6	15	2	5/6	17
	5	9	18	4	8/9	18
	6	9	15	5	25/27	19
	9	9	10	18	80/81	Comb

*The constant $E = tL/rk$ is used in this analysis.
†For the plan classification number see Clatworthy (1973).
‡"Comb" indicates plans constructed by taking all possible combinations of t treatments in groups (block size) of k.

block. In Balanced Incomplete Block Designs, treatments are assigned to blocks that lead to equal precision in the estimation of differences between treatments.

If Randomized Block and Balanced Incomplete Block Designs do not meet the needs of the experimenter with regard to number of blocks, size of blocks, number of treatments, etc., other kinds of plans are available, for example, *partially* balanced incomplete block designs (Clatworthy 1973), and "chain block" designs, which are useful whenever observations are expensive and the experimental error is small. [See Natrella (1963) and Fleiss (1986) for the structure and details of analysis.]

General-purpose software programs written initially for statisticians employing mainframe computers rapidly perform the computations associated with the incomplete block designs. Only the more advanced personal computer design of experiments software programs offer lists of balanced incomplete block designs and assistance in their analysis. Despite their obvious importance and value, the application of these designs within industry is slight.

LATIN SQUARE DESIGNS

A Latin square design (or a Youden square plan, described later) is useful when it is necessary to investigate the effects of different levels of a studied factor while simultaneously allowing for two specific sources of variability or nonhomogeneity, i.e., two different *blocking* variables. Such designs were originally applied in agricultural experimentation when the sources of nonhomogeneity in fertility were simply the two directions on the field, and the "square" was literally a square plot of ground. Its usage has been extended to many other applications in which there are two sources of nonhomogeneity (two blocking variables) that may affect experimental results—for example, machines and positions or operators and days. The studied variable, the experimental treatment, is then associated with the two blocking variables in a prescribed fashion. The use of Latin squares is restricted by two conditions:

1. The numbers of rows, columns, and treatments must all be equal.
2. There must be *no* interactions between the row, the column, and the studied factors (see Factorial Experiments—General, for discussion of interaction).

As an example of a Latin square, suppose we wish to compare four materials with regard to their wearing qualities. Suppose further that we have a wear-testing machine that can handle four samples simultaneously. The two blocking variables might be the variations from run to run and the variations among the four positions on the wear machine. A 4×4 Latin square will allow for both sources of homogeneity. The Latin square plan is shown in Table 47.12 (the four materials are labeled A, B, C, and D). Note that every letter (treatment) appears once in every row and once in every column. Examples of Latin squares from size 3×3 to 7×7 are given in Table 47.13.

Strictly speaking, every time we use a Latin square we should choose a square at random from the set of all possible squares of its size. The tables of Fisher and Yates (1964) give complete collections of all the squares from 3×3 up to 12×12. Once a given square is chosen, permute the

TABLE 47.12 A 4×4 Latin Square

	Position number			
Run	1	2	3	4
1	*A*	*B*	*C*	*D*
2	*B*	*C*	*D*	*A*
3	*C*	*D*	*A*	*B*
4	*D*	*A*	*B*	*C*

TABLE 47.13 Selected Latin Squares

	4 × 4			
3 × 3	1	2	3	4
A B C *B C A* *C A B*	*A B C D* *B A D C* *C D B A* *D C A B*	*A B C D* *B C D A* *C D A B* *D A B C*	*A B C D* *B D A C* *C A D B* *D C B A*	*A B C D* *B A D C* *C D A B* *D C B A*

5 × 5	6 × 6	7 × 7
A B C D E *B A E C D* *C D A E B* *D E B A C* *E C D B A*	*A B C D E F* *B F D C A E* *C D E F B A* *D A F E C B* *E C A B F D* *F E B A D C*	*A B C D E F G* *B C D E F G A* *C D E F G A B* *D E F G A B C* *E F G A B C D* *F G A B C D E* *G A B C D E F*

columns at random, permute the rows at random, and assign the letters randomly to the treatments to provide a completely randomized design.

The analysis of the Latin square design is discussed in most textbooks and is available on most design of experiments personal computer software programs. The analysis of variance table is a simple extension of the randomized block table. Plots of the treatment averages and their reference distribution and/or multiple comparison hypothesis tests are the major objective of the analysis. Plots of the block averages and estimates of block effects are always informative, but since blocks cannot be controlled by the experimenter, their importance is only tangential to the analysis of the treatments. The reader is warned, firmly, that the mathematical model underlying the analysis of a Latin square (or any of its associated designs, the Graeco-Latin and the hyper Graeco-Latin square) assumes that *no interactions* exist between rows, columns, or any of the treatment classifications. Failure to meet this "no interactions" requirement leads to biased estimates of the treatment effects and the row and column (block) effects, and also biases the estimate of the variance σ^2. Unbiased estimates of σ^2 can be obtained by repeating the entire design or by partial replication [see Youden and Hunter (1955)].

The Latin square is *not* a factorial design, i.e., a design that allows for interactions between the separate factors composing the design. If there are interactions likely, the experimenter is advised to use a factorial or fractional factorial design and associated mode of data analysis. The 3×3 Latin square design is sometimes called the L9 orthogonal array. See Hunter (1989) for an example of the dangers that can arise from the misuse of the Latin squares.

YOUDEN SQUARE DESIGNS

The Youden square, like the Latin square, allows for two experimental sources of inhomogeneity. The conditions for the use of the Youden square, however, are less restrictive than those for the Latin square. The use of Latin square plans is restricted by the fact that the number of rows, columns, and treatments must all be the same. Youden squares have the same number of columns and treatments, but a fairly wide choice in the number of rows is possible. We use the following notation:

t = number of treatments to be compared

b = number of levels of one blocking variable (columns)

k = number of levels of another blocking variable (rows)

r = number of replications of each treatment

L = number of times that two treatments occur in the same block

In a Youden square, $t = b$ and $k = r$.

Some Youden square plans are given in Table 47.14. The analysis of the Youden squares must be carefully handled; in particular, the treatment averages must be adjusted for the rows in which they appear *before* they can be compared. Further, the standard error of the adjusted averages requires special computation. Reference should be made to the textbooks listed at the end of this section for the numerical details.

PLANNING INTERLABORATORY TESTS

We present here only a few simple techniques found useful in the planning and analysis of interlaboratory (round-robin) tests. The very early article by Wernimont (1951) remains an excellent introduction to the general problem. Other early contributors to the field are Youden (1967) and Mandel (1964), both members of the early National Bureau of Standards. An overall view of the importance of interlaboratory comparisons can be found in Hunter (1980). The best source of detailed information on round-robin procedures can be found in the publications of committee E11 of the American Society for Testing and Materials (ASTM). The text by Moen, Nolan, and Provost (1991) has several worked examples including a graphical method of analysis due to Snee (1983). Repeatability and Reproducibility (R&R) studies are often part of an interlaboratory testing program; see Barrentine (1991), Montgomery (1991), and Automotive Industry Action Group (1990).

A Rank Sum Test for Laboratories. In almost any set of interlaboratory test data, some of the reported results fall so far out from the main body of results that there is a real question as to whether these data should be omitted in order to avoid distortion of the true picture. It is always a difficult problem to decide whether or not outlying results should be screened. One does not wish to discard a laboratory's results without good reason; on the other hand, if a laboratory is careless or not competent, one does not wish to "punish" the test method. A ranking test for laboratories due to

TABLE 47.14 Youden Square Arrangements

$t = 3$	$t = 4$		$t = 5$
A B	*A B C*		*A B C D*
B C	*B A D*		*B A E C*
C A	*C D B*		*C D A E*
	D C A		*D E B A*
			E C D B
$t = 6$	$t = 7$	or	$t = 7$
A B C D E	*A B D*		*A C D E*
B F D C A	*B C E*		*B D E F*
C D E F B	*C D F*		*C E F G*
D A F E C	*D E G*		*D F G A*
E C A B F	*E F A*		*E G A B*
F E B A D	*F G B*		*F A B C*
	G A C		*G B C D*

Youden (1963) is described here. This is only one of several nonparametric ranking procedures that may be of interest to the reader. Excellent references on these nonparametric approaches are the texts by Hollander and Wolfe (1973), Conover (1980), Gibbons and Chakraborti (1992), and Iman (1994).

An interlaboratory test usually involves sending several materials containing some particular chemical element or compound or possessing some physical quality to each of several laboratories. The ranking test for laboratories uses the recorded measured responses of the materials to rank the laboratories. The data from the interlaboratory test are summarized in a two-way table with materials as rows and laboratories as columns (or vice versa).

For each material, the laboratory having the largest result is given rank 1, the next largest rank 2, etc. (Tied values are treated as is usual in ranking procedures, each tied value being given the average of those ranks that would have been assigned if the values had differed.)

For each laboratory, the assigned ranks are summed over all materials. A laboratory that is consistently high in its ability to measure the response will show a lower rank sum, and a laboratory that is consistently low will show a higher rank sum than the average or expected rank sum. The question is whether such rank sums are excessively high or excessively low. To decide this, tables have been provided (see Table 47.15).

A Ruggedness Test for Use by the Initiating Laboratory. Very often a test method is judged to have acceptable precision by the original laboratory, but when the test is performed by several laboratories, the results are disappointing. The reason is usually that the original laboratory has carefully controlled conditions and equipment and that the operating conditions in other laboratories are slightly different. (There are always slight deviations, which are permissible within the instructions contained in the standard procedure for the test method.) Youden (1967) proposed that the initiating laboratory investigate the effects of such deviations by deliberately introducing small variations in

TABLE 47.15 Approximate 5% Limits for Ranking Scores

Number of laboratories participating	Numbers of materials									
	3	4	5	6	7	8	9	10	11	12
3	· · ·	4	5	7	8	10	12	13	15	17
	· · ·	12	15	17	20	22	24	27	29	31
4	· · ·	4	6	8	10	12	14	16	18	20
	· · ·	16	19	22	25	28	31	34	37	40
5	· · ·	5	7	9	11	13	16	18	21	23
	· · ·	19	23	27	31	35	38	42	45	49
6	3	5	7	10	12	15	18	21	23	26
	18	23	28	32	37	41	45	49	54	58
7	3	5	8	11	14	17	20	23	26	29
	21	27	32	37	42	47	52	57	62	67
8	3	6	9	12	15	18	22	25	29	32
	24	30	36	42	48	54	59	65	70	76
9	3	6	9	13	16	20	24	27	31	35
	27	34	41	47	54	60	73	79	85	91
10	4	7	10	14	17	21	26	30	34	38
	29	37	45	52	60	67	73	80	87	94
11	4	7	11	15	19	23	27	32	36	41
	32	41	49	57	65	73	81	88	96	103
12	4	7	11	15	20	24	29	34	39	44
	35	45	54	63	71	80	88	96	104	112

Note: Let L laboratories test each of M materials. Assign ranks 1 to L for each material. Sum the ranks to get the score for each laboratory. The mean score is $M(L + 1)/2$. The entries are lower and upper limits that are included in the approximate 5% critical region.

the method, a "ruggedness test," so as to be prepared for the variations resulting when the test is used by other laboratories. In order to minimize the extra work required for the original laboratory, he proposed that the Plackett-Burman designs for 7, 11, 15 factors be used to detect such effects. If significant effects result from such variations of conditions in a single laboratory, the method needs further refinement before interlaboratory tests are run. (See Thomas and Kiwanga 1993.)

Youden Two-Sample Plan. A simple plan to investigate the performance of laboratories and of the test procedure itself was suggested by Youden (1959) and reprinted in Ku (1969). Samples of two materials (*A* and *B*) are sent to each laboratory in the program. The two materials should be similar in kind and in the value of the property to be measured. The laboratories should have the same internal precision. The pairs of results are used to plot a graph on which the *x* and *y* scales are equal and each laboratory is represented by one point. A laboratory's result on sample *A* is the *x* coordinate and its result on sample *B* is the *y* coordinate of that point. There will be as many points as there are laboratories. For graphical diagnosis, a vertical line is drawn through the median of all points in the *x* direction and a horizontal line through the median of all points in the *y* direction. The lines could be drawn through the *x* and *y* averages just as well, but the medians are convenient for quick graphical analysis.

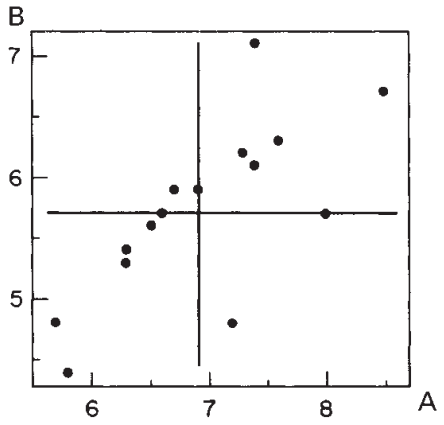


FIGURE 47.4 Percent phthalic anhydride in two paint samples—Youden plot showing systematic differences.

Individual points that are very far removed from the main body of the results indicate laboratories that should probably be screened from the analysis. The two intersecting median lines divide the space into four quadrants, and the first (and often revealing) step in the analysis is to look at the distribution of points among the quadrants. If only random errors of measurement were operating, there would be a circular scatter of points with roughly equal numbers in each quadrant. The plots of most real-life interlaboratory data, however, show concentrations in the upper right and lower left quadrants (see Figure 47.4). If a laboratory is high on both samples, its point will lie in the upper right; if a laboratory is low on both samples, its point will lie in the lower left. Being high (or low) on both samples is an indication that a laboratory has somehow put its own stamp on the procedure, i.e., that there are systematic differences between the laboratories. Where these systematic differences exist, the points will tend to lie along a long, narrow ellipse. Assuming that the two materials are similar in kind and in value of the property measured, as prescribed, and that the scatter in results for sample *A* does turn out to be approximately the same as the scatter for sample *B,* we can calculate an estimate of the standard deviation of a single result as follows:

1. Calculate the "signed differences" $d = A - B$ for each laboratory; that is, compute the difference and keep the sign (for the *i*th laboratory $d_i = A_i - B_i$).
2. Calculate $\bar{d}$, the algebraic average of the *d*'s.
3. Calculate $d'_i = d_i - \bar{d}$.
4. Take the absolute d' values and calculate their average; that is, drop the signs before averaging.
5. Multiply this value by 0.886 to get an estimate *s* of the standard deviation of a single result. (The value 0.886 is $1/d_2$, Appendix II, Table A, for $n = 2$.)

A circle can now be drawn that is expected to contain any stated percentage of the points. The circle is centered at the median point and its radius (for the stated percentage to be contained within it) is obtained by multiplying *s* (from Step 5) by the factor given in Table 47.16.

TABLE 47.16 Radius of Circle on Youden Plot in Terms of Multiples of the Standard Deviation

Percent of points within circle	Multiple of the standard deviation
90	2.146
95	2.448
99	3.035

Points lying outside the circle usually indicate laboratories with systematic differences. Further deductions are possible from such plots (see Youden, 1959); they have been used in a wide variety of applications, including chemical and engineering tests and standards comparisons.

NESTED (COMPONENTS OF VARIANCE) DESIGNS

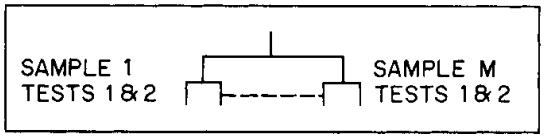


FIGURE 47.5 A two-stage balanced nested design. (*Reprinted with permission from Bainbridge 1965.*)

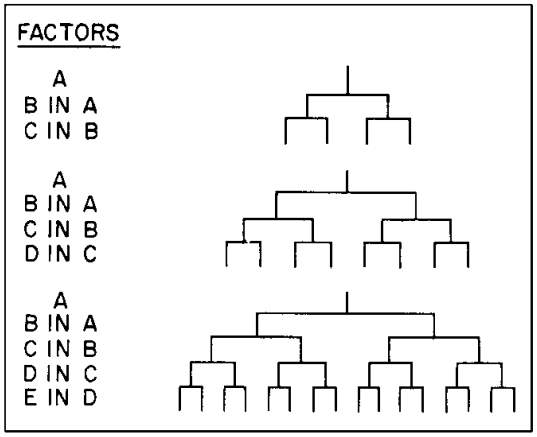


FIGURE 47.6 Balanced nested designs for three, four, and five factors. (*Reprinted with permission from Bainbridge 1965.*)

Most experimental designs are primarily intended to provide estimates of the means and differences or other comparisons between the means of experimental treatments. However, in investigations associated with interlaboratory comparisons, or the repeatability and reproducibility of measuring instruments, sources of variability become the studied factors, and the primary knowledge sought concerns the relative importance of these sources. Such investigations are called "components of variance" studies. Designs intended to provide estimates of the components of variance arising from various sources are called "nested" or "hierarchical" designs. A simple nested design is shown in Figure 47.5 wherein M samples are taken and then within each sample duplicate tests are made. The variance of all the observations is then partitioned into two components: that assignable uniquely to the samples and that uniquely to the duplicates. This is a two-stage balanced nested design. Three or more classifications can be nested, as illustrated in Figure. 47.6.

The analysis of nested designs and estimation of components of variance is based upon a "random effects" model for the observations in which, except for a single constant term, all elements are random variables. (See further discussion under the sections One-Way Analysis of Variance—Models and Split-Plot Factorial Experiments.) Needed are the expected values of the mean squares obtained from an analysis of variance table; see Snee (1974). The designs and their analysis is described in most textbooks on experimental design [for example, Dunn and Clark (1987); Neter, Wasserman, and Kutner (1990); Hogg and Ledolter (1992); Montgomery (1991); and Burdick and Graybill (1992)]. An example of the estimation of components of variance that does not evoke the analysis of variance is given in Box, Hunter, and Hunter (1978). A Bayesian method for estimating variance components is given in Box and Tiao

(1973). An entire textbook devoted to the subject is Searle, Casella, and McCulloch (1992). Most design of experiments software programs will provide the expected values of the mean squares needed to determine the components of variance for the simple balanced designs.

In Figure 47.6, at each stage only two subunits of each unit are taken, but the total number of tests multiplies rapidly as the number of stages increases. Because of the rapidly increasing total number of tests, only a few units are usually used at the top levels. In other words, balanced nested designs tend to provide too little information on the upper levels (the initial stages, or factors *A* and *B*) and often provide more than enough information at the bottom levels (factors *E,* for example). Bainbridge (1965) has considered alternative "unbalanced" nested designs with a fixed number of tests. He prefers a design, which he calls a "staggered nested design" that is easy to administer and provides about the same number of degrees of freedom for each factor. Bainbridge shows staggered nested designs for three, four, five, and six factors (see Figure 47.7 for the designs and their analysis).

PLANNING THE SIZE OF THE EXPERIMENT

Methods for determining the number of observations required for estimating the mean and variance with certain precision, or for comparing two sets of data with regard to mean and variance with certain risks of error, are given in Section 44, Basic Statistical Methods. A method for determining the number of observations required for comparing several groups is given here.

For example, the analysis of variance *F* test (see Completely Randomized Design: One Factor, *k* Levels) is designed to test the hypothesis that all group means are the same, i.e., $\eta_1 = \eta_2 \ldots \eta_k = \eta$. The corresponding averages $\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_k$ computed from the recorded data will, of course, be different. The outcome of the test of hypothesis depends on the significance level α at which the test is performed, the true variability of individual observations, the number of observations per average, and the size of the true difference (if any) between group means. When planning experiments, if there are no restrictions on the number of observations that can be made, one should specify the size of those differences in means that are considered important from a practical standpoint. When the significance level at which the test is to be made is also specified, existing tables or charts can be used to determine the necessary sample size (number of observations per average) for achieving a stated probability $(1 - \beta)$ of detecting differences between the means of the required size. To use such tables, we compute a quantity

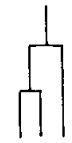
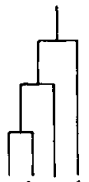
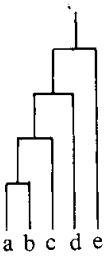
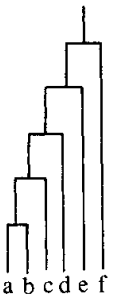
$$\phi^2 = \frac{n \sum_i (\eta_i - \eta)^2}{k\sigma^2} = \frac{n\Sigma\delta_i^2}{k\sigma^2}$$

where n = number of observations per group (to be determined)
k = number of groups
σ^2 = true value of within-group variance (assumed same for all groups; can be estimated from previous similar work)
η_i = mean for *i*th group
η = grand mean

Let $(\eta_i - \eta) = \delta_i$. The sum of the δ_i values must equal zero.

Appendix II, Table DD gives values ϕ^2 for $\alpha = 0.01$ and $\beta = 0.2$, $(1 - \beta = 0.80)$, and DF_1 and DF_2 degrees of freedom. In the simple case used in the example, $DF_1 = k - 1$ and $DF_2 = k(n - 1)$. Other charts and tables are available in slightly different form and for additional values of α and β. See, for example, Dixon and Massey (1969), Owens (1962), or Odeh and Fox (1991).

Example. Consider the experiment shown in Table 47.4. Suppose that another experiment is to be run and that we wish to determine beforehand how many briquettes to test using each method in

Sources of Variance	Sums of Squares	Degrees of Freedom	Expectations of Mean Squares	Format of A-Units
A	(5) - CF	m-1	$\sigma_c^2 + 1^2/_3\,\sigma_b^2 + 3\sigma_a^2$	
B in A	(3) + (4) - (5)	m	$\sigma_c^2 + 1^1/_3\,\sigma_b^2$	
C in B	(1) + (2) - (3)	m	σ_c^2	
Total	(1) + (2) + (4) - CF	3m-1	4A Three Factors	a b c
A	(7) - CF	m-1	$\sigma_d^2 + 1^1/_2\,\sigma_c^2 + 2^1/_2\,\sigma_b^2 + 4\sigma_a^2$	
B in A	(5) + (6) - (7)	m	$\sigma_d^2 + 1^1/_6\,\sigma_c^2 + 1^1/_2\,\sigma_b^2$	
C in B	(3) + (4) - (5)	m	$\sigma_d^2 + 1^1/_3\,\sigma_c^2$	
D in C	(1) + (2) - (3)	m	σ_d^2	
Total	(1) + (2) + (4) + (6) - CF	4m-1	4B Four Factors	a b c d
A	(9) - CF	m-1	$\sigma_e^2 + 1^2/_5\,\sigma_d^2 + 2^1/_5\,\sigma_c^2 + 3^2/_5\,\sigma_b^2 + 5\sigma_a^2$	
B in A	(7) + (8) - (9)	m	$\sigma_e^2 + 1^1/_{10}\,\sigma_d^2 + 1^3/_{10}\,\sigma_c^2 + 1^3/_5\,\sigma_b^2$	
C in B	(5) + (6) - (7)	m	$\sigma_e^2 + 1^1/_6\,\sigma_d^2 + 1^1/_2\,\sigma_c^2$	
D in C	(3) + (4) - (5)	m	$\sigma_e^2 + 1^1/_3\,\sigma_d^2$	
E in D	(1) + (2) - (3)	m	σ_e^2	
Total	(1) + (2) + (4) + (6) + (8) - CF	5m-1	4C Five Factors	a b c d e
A	(11) - CF	m-1	$\sigma_f^2 + 1^1/_3\,\sigma_e^2 + 2\sigma_d^2 + 3\sigma_c^2 + 4^1/_3\,\sigma_b^2 + 6\sigma_a^2$	
B in A	(9) + (10) - (11)	m	$\sigma_f^2 + 1^1/_{15}\,\sigma_e^2 + 1^1/_5\,\sigma_d^2 + 1^2/_5\,\sigma_c^2 + 1^2/_3\,\sigma_b^2$	
C in B	(7) + (8) - (9)	m	$\sigma_f^2 + 1^1/_{10}\,\sigma_e^2 + 1^3/_{10}\,\sigma_d^2 + 1^3/_5\,\sigma_c^2$	
D in C	(5) + (6) - (7)	m	$\sigma_f^2 + 1^1/_6\,\sigma_e^2 + 1^1/_2\,\sigma_d^2$	
E in D	(3) + (4) - (5)	m	$\sigma_f^2 + 1^1/_3\,\sigma_e^2$	
F in E	(1) + (2) - (3)	m	σ_f^2	
Total	(1) + (2) + (4) + (6) + (8) + (10) - CF	6m-1	4D Six Factors	a b c d e f

TOTALS NEEDED TO GET SUMS OF SQUARES

$$(1) = \Sigma a^2 \qquad (5) = \frac{\Sigma\,(a+b+c)^2}{3} \qquad (9) = \frac{\Sigma\,(a+b+c+d+e)^2}{5}$$

$$(2) = \Sigma b^2 \qquad (6) = \Sigma d^2 \qquad (10) = \Sigma f^2$$

$$(3) = \frac{\Sigma\,(a+b)^2}{2} \qquad (7) = \frac{\Sigma\,(a+b+c+d)^2}{4} \qquad (11) = \frac{\Sigma\,(a+b+c+d+e+f)^2}{6}$$

$$(4) = \Sigma c^2 \qquad (8) = \Sigma e^2 \qquad CF = \frac{(\text{Grand Total})^2}{\text{Total No. of Tests}}$$

FIGURE 47.7 Staggered nested designs for three, four, five, and six factors. (*Reprinted with permission from Bainbridge 1965.*)

order to achieve a certain discrimination between the means for the three methods. If the statistical test is to be done at the $\alpha=0.01$ level and if we want the probability of detecting the postulated differences to be at least 0.8, we can use Appendix II, Table DD. Assume $\sigma^2=545$, an estimate of the variance determined from the previous experiment. Suppose the following differences between the means are considered practically important:

$$\delta_1 = \eta_1 - \eta = -30$$

$$\delta_2 = \eta_2 - \eta = +20$$

$$\delta_3 = \eta_3 - \eta = +10$$

(Obviously, many different values for the δ's will yield the same value for $\Sigma\delta^2$ and therefore the same ϕ^2.)

The δ's chosen should be meaningful for each experimental situation. Here we have postulated three particular differences; in other situations the pattern of the differences might take on special meanings. For example, if the groups were increasing levels of a quantitative variable such as temperature, a meaningful pattern for the δ's might be a constant change in mean from one level to the next higher one. (Remember that the δ_i's must sum to zero.)

$$\phi^2 = \frac{n\Sigma\delta_i^2}{k\sigma^2}$$

$$\phi^2 = \frac{n(1400)}{3(545)} = \frac{1400n}{1635} = 0.86n$$

$$DF_1 = k - 1 = 3 - 1 = 2$$

$$DF_2 = k(n - 1) = 3n - 3$$

Using Appendix II, Table DD, we must find two values of n, one that gives ϕ^2 larger than required and one that gives a smaller value than required:

n	$DF_2 = 3n - 3$	Tabled ϕ^2	Desired $\phi^2 = 0.86n$
7	18	6.05	6.02
8	21	5.83	6.88

The "tabled ϕ^2" for $n = 8$ was obtained by linear interpolation. The solution lies between $n = 7$ and $n = 8$, and we take the larger n. Eight observations per group will give us an 80 percent chance of detecting the postulated differences when we do an F test at the $\alpha = 0.01$ level.

This method may be used for multifactor experiments provided the proper values for DF_1 and DF_2 are used. It is used when the purpose of the experiment is to compare group averages, and it works for any number of groups provided the number of observations per group is large enough. In this case and in the case described below, equal numbers of observations should be taken in each group.

For another kind of experiment, in which the purpose is to compare the between-group variance with the within-group variance (see discussion of Model II, Random Effects Model, under One-Way Analysis of Variance—Models), a *minimum* number of *groups* is required to achieve desired discrimination in terms of the relative variability. For example, see Table 47.17, where α and β are the risks of the two kinds of error, δ_0 is an "acceptably small" value of the ratio σ_b/σ_w (large enough to achieve a significant result), and δ_1 is an unacceptably large value for σ_b/σ_w.

TABLE 47.17 Minimum Number of Groups—Random Effects Model

$\alpha = \beta = 0.05$		$\alpha = \beta = 0.01$	
δ_1/δ_0	Minimum number of groups	δ_1/δ_0	Minimum number of groups
1.5	35	1.5	68
2.0	14	2.0	25
2.5	9	2.5	16
3.0	7	3.0	12

Useful discussions on determining the number of observations are given in the texts by Cochran and Cox (1957), Cox (1958), and Odeh and Fox (1991). Extensive tables are given in the papers by Kastenbaum, Hoel, and Bowman (1970).

FACTORIAL EXPERIMENTS—GENERAL

Factorial designs are most frequently employed in engineering and manufacturing experiments. In a factorial experiment several factors are controlled at two or more levels, and their effects upon some response are investigated. The experimental plan consists of taking an observation at each of all possible combinations of levels that can be formed from the different factors. Each different combination of factor levels is called a "treatment combination."

Suppose that an experimenter is interested in investigating the effect of two factors, amperage (current) level and force, upon the response y, the measured resistivity of silicon wafers. In the past, one common experimental approach has been the so-called one-factor-at-a-time approach. This experimental strategy studies the effect of first varying amperage levels at some constant force and then applying different force levels at some constant level of amperage. The two factors would thus be varied one at a time with all other conceivable factors held as constant as possible. The results of such an experiment are fragmentary in the sense that we learn about the effect of different amperage levels only at one force level and the effect of different force levels at only one amperage level. The effects of one factor are conditional on the chosen level of the second factor. The measured resistivity of the wafer at different current levels may, of course, be different when a different force level has been chosen. Similarly, any observed relation of resistivity to force level might be quite different at other amperage levels. In statistical language, there may be an "interaction effect" between the two factors over the range of interest, and the one-at-a-time procedure does not enable the experimenter to detect the interaction.

In a factorial experiment, the levels of each factor are chosen, and a measurement is made at each of all possible combinations of levels of the factors. Suppose that five levels of amperage and four levels of force are chosen. There would thus be 20 possible combinations of amperage and force, and the factorial experiment would consist of 20 trials. In this example, the term "level" is used in connection with quantitative factors, but the same term is also used when the factors are qualitative.

In the analysis of factorial experiments, one speaks of "main effects" and "interaction effects" (or simply "interactions"). Estimated main effects of a given factor are always functions of the average yield response at the various levels of the factor. When a factor has two levels, the estimated main effect is the difference between the average responses at the two levels, i.e., the averages computed over all levels of the other factors. In the case in which the factor has more than two levels, there are several main effect components (linear, quadratic, cubic, etc.), the number of estimable main effect components being one less than the number of levels. Other comparisons, called treatment "contrasts," are possible. If the difference in the expected response between two levels of factor A remains constant over the levels of factor B (except for experimental error), there is no interaction between

A and *B*; that is, the *AB* interaction is zero. Figure 47.8 shows two examples of response, or yield, curves; one example shows the presence of an interaction and the other shows no interaction. If there are two levels each of the factors *A* and *B,* then the *AB* interaction (neglecting experimental error) is the difference in the average yields of *A* at the second level of *B* minus the difference in the average yields of *A* at the first level of *B.* If there are more than two levels of either *A* or *B,* then the *AB* interaction can be composed of more than one component. If we have a levels of the factor *A* and b levels of the factor *B,* then the *AB* interaction has $(a-1)(b-1)$ independent components. A two-factor interaction (e.g., *AB*) is also called a "second-order" effect or "coupled" effect.

For factorial experiments with three or more factors, interactions can also be defined. For example, the *ABC* interaction is the interaction between the factor *C* and the *AB* interaction (or, equivalently, between the factor *B* and the *AC* interaction or between *A* and the *BC* interaction). A three-factor interaction (e.g., *ABC*) is a "third-order" effect.

FACTORIAL EXPERIMENT WITH TWO FACTORS

A two-factor experiment is the simplest kind of multifactor experiment; i.e., all possible combinations of the levels of the two factors are run. For example, measurements of the response resistivity of a silicon wafer are usually made at a standard amperage level while using 150 g of force. Let us consider an investigation in progress to see what happens when other values of force and amperage are employed. Four values of force are to be investigated (25, 50, 100, 150 g), along with five levels of amperage (levels 1, 2, 3, 4, and 5, where level 3 is the standard level). An experimental trial is made at each of the $4\times5=20$ possible combinations. The data can be displayed in a two-way array, as in Table 47.18. This is an unreplicated two-factor multilevel factorial experiment. Some textbooks will describe it as having two "crossed" factors.

Analysis. The data for this 4×5 factorial are displayed in Table 47.18. The first stage of the analysis is to compute, plot, and review both the column and row averages shown in Table 47.19.

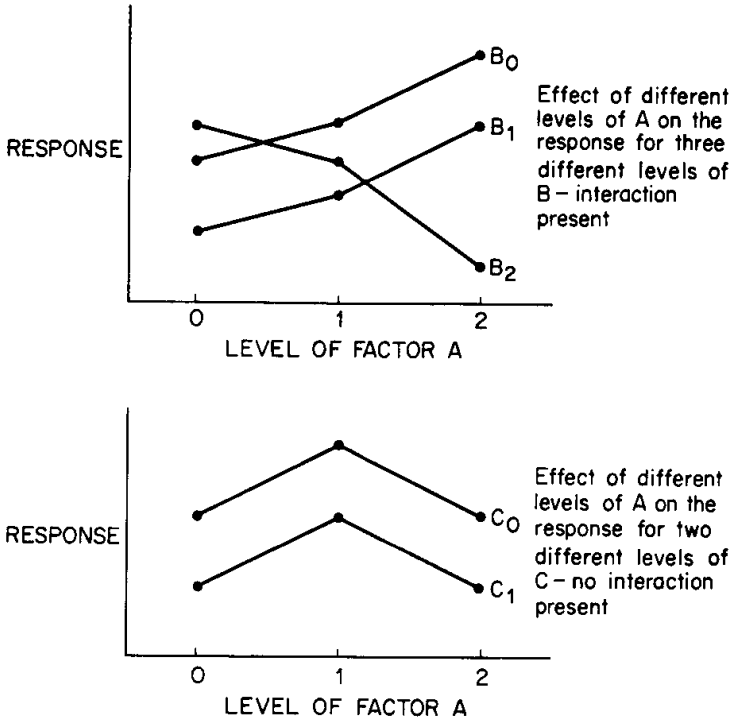


FIGURE 47.8 Response curves showing presence or absence of interaction.

TABLE 47.18 Resistivity Measurements

	Current level					
Force, g	1	2	3	4	5	Row totals R_i
25	11.84	11.83	11.84	11.81	11.96	$R_1 = 59.28$
50	11.84	11.88	11.88	11.87	11.90	$R_2 = 59.37$
100	11.77	11.80	11.80	11.81	11.88	$R_3 = 59.06$
150	11.79	11.80	11.80	11.80	11.87	$R_4 = 59.06$
Column totals C_j	47.24	47.31	47.32	47.29	47.61	$T = 236.77$

TABLE 47.19 Table of Averages and Effects

Row number	Row average	Row effect
1	11.8560	0.0175
2	11.8740	0.0355
3	11.8120	−0.0265
4	11.8120	−0.0265
Column number	**Column average**	**Column effect**
1	11.8100	−0.0285
2	11.8275	−0.0110
3	11.8300	−0.0085
4	11.8225	−0.0160
5	11.9025	0.0640

Plots of the response versus the factor levels are always useful; in this case they would point up the noticeably higher resistivity values at amperage level 5 and the apparent changes in the resistivity with increasing force, particularly at the outer values of amperage. When the factors are quantitative and the levels equally spaced, there are simple methods to check on various types of trends (linear, quadratic, etc.) in the response measurements as a function of varying levels of a factor (see Hicks 1982).

The analysis of variance can be used to test two hypotheses: (1) the mean resistivity at all levels of force is the same and (2) the mean resistivity at all levels of amperage is the same.

The construction of the analysis of variance table would proceed as follows:

$$r = \text{number of rows} = 4$$

$$c = \text{number of columns} = 5$$

$$T = \text{grand total} = \sum_j C_j = \sum_i R_i = 236.77$$

$$N = \text{total number of observations} = r \times c = 20$$

$$C = \text{the correction factor} = \frac{T^2}{N} = 2803.001645$$

$$\text{SSR} = \text{row sum of squares} = \frac{\sum_i R_i^2}{c} - C = 0.014855$$

$$\text{SSC} = \text{column sum of squares} = \frac{\sum_j C_j^2}{r} - C = 0.021430$$

$$\begin{aligned}
\text{TSS} &= \text{total corrected sum of squares} \\
&= (\text{all observations squared}) - C \\
&= 2803.043500 - 2803.001645 = 0.041855 \\
\text{SSE} &= \text{error (or residual) sum of squares} \\
&= \text{TSS} - \text{SSR} - \text{SSC} = 0.005570
\end{aligned}$$

The analysis of variance table for this (unreplicated) two-factor crossed factorial is shown in Table 47.20. Tests of hypotheses that no treatment effects exist are now possible. For rows conclude that statistically significant effects exist between rows if F = (row mean square)/(error mean square) is greater than $F_{1-\alpha}$ for $(r-1)$ and $(r-1)(c-1)$ degrees of freedom (from Appendix II, Table K). Similarly, for columns conclude that statistically significant effects exist if F = (column mean square)/(error mean square) is greater than the tabulated value of F for $(c-1)$ and $(r-1)(c-1)$ degrees of freedom. In this example, both F tests reject the hypothesis that no effects exist.

The analysis of variance table and its associated tests of hypotheses are only one part of the analysis of data from a factorial design. The model postulated for the data assumes that the force and amperage factor effects are additive. Thus there can be no interactions; that is, the effects of force upon resistivity remain unchanged whatever the levels of the second factor. One way to check this assumption is to compute Tukey's one-degree-of-freedom test for nonadditivity (see Snedecor and Cochran 1967 or Box, Hunter, and Hunter 1978). The test for nonadditivity here proved to be nonsignificant.

Table 47.19 has been extended to include the estimated row and column *effects* (an estimated row effect is the row average minus the grand average; an estimated column effect is the column average minus grand average). The row and column effects can then be used to make a two-way table of "residuals" in which the residual for the cell in the *i*th row and *j*th column is equal to the observation in that cell minus the sum of the grand average, the *i*th row effect, and the *j*th column effect. The table of residuals, Table 47.21, is examined for individual large values (indicating a possibly erroneous observation) and for unusual patterns in sign and size (indicating possible interaction effects). The residuals should be plotted in the time order in which the treatment combinations were run; any indication of trends indicative of other factors disturbing the response. The residuals can also be plotted on normal probability paper as a check on the normality assumption.

The discussion thus far assumes that only one determination per cell was made. To obtain a truly valid estimate of the error variance from this experiment, the cell's observations must be replicated. When experiments are replicated (ideally with each cell containing the same number of observations), it is useful to have a table similar to Table 47.18, where now in each cell both the average and the estimate of variance are recorded. Homogeneity of variance tests may be made (see Snedecor and Cochran 1967; Duncan 1974; Dyer and Keating 1980), although it is well to remember that such

TABLE 47.20 Analysis of Variance of Resistivity Measurements Given in Table 47.18

Source of variation	Sum of squares (SS)	Degrees of freedom (DF)	Mean square = SS/DF	F*
Rows (force)	0.014855	$(r-1) = 3$	0.004952	10.67†
Columns (current)	0.021430	$(c-1) = 4$	0.005358	11.54†
Error	0.005570	$(r-1)(c-1) = 12$	0.000464	
Total	0.041855	$(rc-1) = 19$		

*F = mean square (source)/mean square (error).
†Significant at the 1% level.

TABLE 47.21 Residuals, Row and Column Effects, Grand Average

Force, g	Amperage, A					
	Level 1	Level 2	Level 3	Level 4	Level 5	Force effects
25	0.0125	−0.0150	−0.0075	−0.0300	0.0400	0.0175
50	−0.0055	0.0170	0.0145	0.0120	−0.0380	0.0355
100	−0.0135	−0.0010	−0.0035	0.0140	0.0040	−0.0265
150	0.0065	−0.0010	−0.0035	0.0040	−0.0060	−0.0265
Current effects	−0.0285	−0.0110	−0.0085	−0.0160	0.0640	11. 8385 = $\bar{y}$

tests require near exact normality if they are to be useful. The analysis of variance assumes these variance estimates to be homogeneous and pools them.

A plot of each cell average versus cell estimate of variance (or standard deviation) is often revealing. Individual outlying points or a pattern of dependence of variability on average value should be looked for. (In the latter case, the need for a transformation of the data should be considered; see Box, Hunter, and Hunter 1978.)

In this experiment, replicate measurements were made in each cell of the table since the investigator was interested in finding out about possible interactions and whether the variance of the resistivity measurements would be constant at the extreme values of the factors. In this original experiment, no significant interactions were found, nor was the variance nonhomogeneous.

The analysis of variance of a replicated two-factor crossed factorial design experiment is easily modified from that of the unreplicated one. The procedure is to calculate the following quantities and insert them in Table 47.22:

r = number of rows

c = number of columns

k = number of determinations per cell

N = total number of observations = krc

T = grand total

$C = T^2/N$

$$\text{SSR} = \text{row sum of squares} = \frac{\sum_i R_i^2}{kc} - C$$

$$\text{SSC} = \text{column sum of squares} = \frac{\sum_j C_j^2}{kr} - C$$

$$\text{SSI} = \text{Interaction sum of squares} = \frac{\sum_{ij} (\text{cell}_{ij}\ \text{total})^2}{k} - \text{SSR} - \text{SSC} - C$$

$$\text{TSS} = \text{total sum of squares} = \Sigma y^2 - C$$

$$\text{SSE} = \text{error sum of squares} = \text{TSS} - \text{SSR} - \text{SSC} - \text{SSI}$$

TABLE 47.22 Analysis of Variance Table for Two-Factor Factorial (k Replicates per Cell)

Source of variation	Sum of squares	Degrees of freedom	Mean square	F
Rows	SSR	$r - 1$	$\text{SSR}/(r - 1) = \text{MSR}$	MSR/MSE
Columns	SSC	$c - 1$	$\text{SSR}/(c - 1) = \text{MSC}$	MSC/MSE
Interaction	SSI	$(r - 1)(c - 1)$	$\text{SSI}/(r - 1)(c - 1) = \text{MSI}$	MSI/MSE
Error	SSE	$rc(k - 1)$	$\text{SSE}/rc(k - 1) = \text{MSE}$	
Total	SST	$krc - 1$		

The above instructions will fill in all the cells in the "Sum of squares" column of the analysis of variance table (Table 47.22), but the similarities to Table 47.5 should be noted. The value of the mean square error (MSE) in Table 47.22 can also be obtained by pooling the ($r \times c$) estimates of the within-cell variance.

SPLIT-PLOT FACTORIAL EXPERIMENTS

In general a factorial design is a program of experiments consisting of all possible combinations of levels (versions) of k different factors (variables). There can be m_1 levels of factor 1 combined with m_2 levels of factor 2,…, combined with m_k levels of factor k to give a total of $N = m_1 \times m_2 \times \ldots \times m_k$ experimental trials.

Care must be taken that the N experimental trials composing a factorial design are all run in a random sequence. Consider for example an experimental program for studying temperature and pressure each at three levels along with four different versions of catalyst, that is, a $3 \times 3 \times 4$ factorial design in $N = 36$ trials. To be a standard factorial all 36 runs would be performed in random order. However, it is very likely that the experimental design will be run in a split-plot (nested) arrangement; that is, the nine temperature-pressure runs will be randomly performed for each catalyst separately. The analysis of a split-plot factorial design, most particularly the computation of correct confidence intervals and tests of hypotheses concerning treatment means, usually requires the attention of a professional statistician. In this handbook only fully randomized designs are discussed.

Split-plot designs are alike in structure to the "nested" designs discussed earlier. Commonly, the designs are called "split-plot" when the objective is the study of treatment means and called "nested" when the objective of the experimenter is to estimate components of variance. Statisticians, and most computer software programs, distinguish between these two forms of experimental design with their different objectives. The factorial split-plot arrangement requires a "Type I" analysis of variance while that for the nested designs requires a "Type II." Occasionally, experimental designs are employed in which estimates of both means and components of variance are required. The subtleties of analysis become important and one should seek the advice of a professional statistician.

FACTORIAL EXPERIMENTS WITH k FACTORS (EACH FACTOR AT TWO LEVELS)

The 2^k factorial designs have widespread industrial applicability. The designs permit the separate estimation of the individual effects and the interaction effects of the k factors in an experimental program in which all k factors are varied simultaneously in a carefully organized pattern of trials.

Symbols. A factorial experiment with k factors, each at two levels, is known as a 2^k factorial experiment. The experiment consists of 2^k trials, one trial at each combination of levels of the fac-

TABLE 47.23 Different Notations for the 2^3 Factorial Design

	Geometric notation			Alternative notation			Japanese notation			Classical notation		
Run no.	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*
1	−	−	−	0	0	0	1	1	1		1	
2	+	−	−	1	0	0	2	1	1		*a*	
3	−	+	−	0	1	0	1	2	1		*b*	
4	+	+	−	1	1	0	2	2	1		*ab*	
5	−	−	+	0	0	1	1	1	2		*c*	
6	+	−	+	1	0	1	2	1	2		*ac*	
7	−	+	+	0	1	1	1	2	2		*bc*	
8	+	+	+	1	1	1	2	2	2		*abc*	

tors. To identify the individual trials, different notations are used, as illustrated in Table 47.23. One convention is to label each factor by a letter (or numeral) and then to denote the two levels (versions) of each factor by a plus (+) and a minus (−) sign. Commonly the minus sign refers to the lower level, the standard conditions, or the absence of the factor. Thus, if there are three factors labeled *A, B,* and *C,* the eight trials comprising the 2^3 factorial design are as shown in Table 47.23. The (+, −) notation is sometimes referred to as "geometric." For example, the eight (±, ±, ±) factor settings for the 2^3 design may be interpreted as giving the (±1, ±1, ±1) coordinates of the eight vertices of a cube. Alternative notations are to employ 0 and 1, respectively, or, following the Japanese tradition earlier established by Taguchi, 1 and 2 for the two versions of each factor. The classical convention is to denote the two versions of each lettered factor by the presence and absence of its corresponding lowercase letter, as is also illustrated in Table 47.23. Here the trial in which all factors are at their "low" level is denoted by a 1. The sequence of trials in Table 47.23 is written in standard or "Yates" order. The trials would, of course, be run in random order.

Example. The data in Table 47.24 are taken from a larger experiment on fire-retardant treatments for fabrics. The excerpted data are intended only to provide an example for demonstrating the technique of analysis. The experiment has four factors, each at two levels, i.e., it is a 2^4 factorial. Note that all factors are qualitative in this experiment. The experimental factors and levels (versions) are

Factors	Levels
A—Fabric	−Sateen
	+Monk's cloth
B—Treatment	−Treatment *x*
	+Treatment *y*
C—Laundering	−Before laundering
	+After one laundering
D—Direction	−Warp
	+Fill

The observations reported in Table 47.24 are inches burned, measured on a standard-sized sample after a flame test. For convenience, alternative design notations representing the treatment combinations appear beside the resulting observation.

Estimation of Main Effects and Interactions. Obtaining the estimates of main effects and interactions from a 2^k factorial design (and 2^{k-p} fractional factorial designs) is available on

TABLE 47.24 Results of Flame Test of Fire-Retardant Treatments (a 2^4 Factorial Experiment)

Treatment combinations					
Geometric notation				Classical notation	Response yield, inches burned
A	*B*	*C*	*D*		
−	−	−	−	1	4.2
+	−	−	−	*a*	3.1
−	+	−	−	*b*	4.5
+	+	−	−	*ab*	2.9
−	−	+	−	*c*	3.9
+	−	+	−	*ac*	2.8
−	+	+	−	*bc*	4.6
+	+	+	−	*abc*	3.2
−	−	−	+	*d*	4.0
+	−	−	+	*ad*	3.0
−	+	−	+	*bd*	5.0
+	+	−	+	*abd*	2.5
−	−	+	+	*cd*	4.0
+	−	+	+	*acd*	2.5
−	+	+	+	*bcd*	5.0
+	+	+	+	*abcd*	2.3

almost all design of experiments software programs. The computations are simple to do by hand but become quickly tedious and hence ideal work for a computer (see Bisgaard 1993a). Further, the associated analysis of variance table is similarly easy and hence widely available. Nevertheless, we include here a description of the hand computations for the interested reader, since they are also easily done on a spread sheet.

The 2^k factorial designs permit the estimation of all k main effects (first-order effects), all $k(k-1)/2$ two-factor interactions, all $k(k-1)(k-2)/3!$ three-factor interactions, etc. Each estimated effect is a statistic of the form $\bar{y}_+ - \bar{y}_-$; that is, it is expressed by the difference between two averages, each containing 2^{k-1} observations. For a 2^4 design the analyst would thus be able to estimate, in addition to the grand average, four main effects, six two-factor interactions, four three-factor interactions, and a single four-factor interaction, giving a total of 16 statistics. Remarkably, all these statistics are "clear" (orthogonal) of one another; that is, the magnitudes and signs of each statistic are in no manner influenced by the magnitude and sign of any other.

The question as to which observations go into which average for each estimated effect is determined from the k columns of + and − signs that together form the experimental design (the design column "vectors"). Additional column vectors of + and − signs are then constructed for each interaction, as illustrated in Table 47.25. For example, the vector of signs labeled *AB* is obtained by algebraically multiplying, for each row, the + or − sign found in column A by the + or − sign found in column *B*.

Table 47.25 also contains the column of observations. To estimate the *AB* interaction effect, all the observations carrying a + sign in the *AB* column are placed in $\bar{y}_+$ and those with a minus sign in $\bar{y}_-$. The estimated *AB* interaction effect ($\bar{y}_+ - \bar{y}_-$) is therefore:

$$\frac{4.2 + 2.9 + \ldots + 2.3}{8} - \frac{3.1 + 4.5 + \ldots + 5.0}{8} = \frac{27.0}{8} - \frac{30.5}{8} = \frac{-3.5}{8} = -0.4375$$

TABLE 47.25 Table of Signs for Calculating Effects for a 2^4 Factorial

A	*B*	*C*	*D*	*AB*	*AC*	*AD*	*BC*	*BD*	*CD*	*ABC*	*ABD*	*ACD*	*BCD*	*ABCD*	Obs.*
−	−	−	−	+	+	+	+	+	+	−	−	−	−	+	4.2
+	−	−	−	−	−	−	+	+	+	+	+	+	−	−	3.1
−	+	−	−	−	+	+	−	−	+	+	+	−	+	−	4.5
+	+	−	−	+	−	−	−	−	+	−	−	+	+	+	2.9
−	−	+	−	+	−	+	−	+	−	+	−	+	+	−	3.9
+	−	+	−	−	+	−	−	+	−	−	+	−	+	+	2.8
−	+	+	−	−	−	+	+	−	−	−	+	+	−	+	4.6
+	+	+	−	+	+	−	+	−	−	+	−	−	−	−	3.2
−	−	−	+	+	+	−	+	−	−	−	+	+	+	−	4.0
+	−	−	+	−	−	+	+	−	−	+	−	−	+	+	3.0
−	+	−	+	−	+	−	−	+	−	+	−	+	−	+	5.0
+	+	−	+	+	−	+	−	+	−	−	+	−	−	−	2.5
−	−	+	+	+	−	−	−	−	+	+	+	−	−	+	4.0
+	−	+	+	−	+	+	−	−	+	−	−	+	−	−	2.5
−	+	+	+	−	−	−	+	+	+	−	−	−	+	−	5.0
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	2.3

*Obs. = observations.

Yates' Algorithm. An alternative and more rapid method for obtaining estimates of main effects and interactions for two-level factorials, called "Yates' algorithm," applies to all two-level factorials and fractional factorials. The first step in Yates' algorithm is to list the observed data in Yates order, as illustrated in Table 47.26. The generation of the values in Table 47.26 proceeds as follows:

1. A two-level factorial with r replicates contains $N = r2^k$ runs. The associated Yates' algorithm table will have $k + 2$ columns, the first of which contains the experimental design, i.e., the 2^k treatment combinations in standard (Yates) order.
2. In column 2, enter the observed yield corresponding to each treatment combination listed in column 1. If the design is replicated, enter the total for each treatment combination.
3. In the top half of column 3 enter, in order, the sums of consecutive *pairs,* of entries in column 2, i.e., the first plus the second, the third plus the fourth, and so on. In the bottom half of column 3 enter, in order, the differences between the same consecutive pairs of entries, i.e., second entry minus first entry, fourth entry minus third entry, etc. Change the sign of the top (first of the pair) and algebraically add.
4. Obtain columns 4, 5, …, k+2, in the same manner as column 3, i.e., by obtaining in each case the sums and differences of the pairs in the preceding column in the manner described in step 3.
5. The entries in the last column (column $k + 2$) are labeled $g(T)$, $g(A)$, $g(B)$, $g(AB)$, etc. The letters in the parentheses correspond to the + signs in the geometric notation. The first value $g(T)$ is divided by N to give the grand average. Estimates of the remaining main effects and interactions are obtained by dividing each $g(\ldots)$ by $N/2$. (*Note:* The remaining steps of this procedure are checks on the computations.)
6. The sum of all the individual responses (column 2) should equal the total given in the first entry of column 6, i.e., $g(T)$ must equal the grand total.
7. The sum of the squares of the quantities in column 2 should equal the sum of the squares of the entries in column $(k + 2)$ divided by 2^k.
8. Each $g(\ldots)$ in the last column equals the sum of observations carrying a + sign minus the sum of observations carrying a − sign when the columns of signs displayed in Table 47.25 are employed.

TABLE 47.26 Yates Method of Analysis Using Data of Table 47.24

A *B* *C* *D*	2	3	4	5	6	Estimated effects
− − − −	4.2	7.3	14.7	29.2	57.5 = $g(T)$	Average = 3.5938
+ − − −	3.1	7.4	14.5	28.3	−12.9 = $g(A)$	A = −1.6125
− + − −	4.5	6.7	14.5	−5.2	2.5 = $g(B)$	B = 0.3125
+ + − −	2.9	7.8	13.8	−7.7	− 3.5 = $g(AB)$	AB = −0.4375
− − + −	3.9	7.0	−2.7	1.2	− 0.9 = $g(C)$	C = −0.1125
+ − + −	2.8	7.5	−2.5	1.3	− 0.5 = $g(AC)$	AC = −0.0625
− + + −	4.6	6.5	−3.5	−0.8	1.3 = $g(BC)$	BC = 0.1625
+ + + −	3.2	7.3	−4.2	−2.7	0.5 = $g(ABC)$	ABC = 0.0625
− − − +	4.0	−1.1	0.1	−0.2	− 0.9 = $g(D)$	D = −0.1125
+ − − +	3.0	−1.6	1.1	−0.7	− 2.5 = $g(AD)$	AD = −0.3125
− + − +	5.0	−1.1	0.5	0.2	0.1 = $g(BD)$	BD = 0.0125
+ + − +	2.5	−1.4	0.8	−0.7	− 1.9 = $g(ABD)$	ABD = −0.2375
− − + +	4.0	−1.0	−0.5	1.0	− 0.5 = $g(CD)$	CD = −0.0625
+ − + +	2.5	−2.5	−0.3	0.3	− 0.9 = $g(ACD)$	ACD = −0.1125
− + + +	5.0	−1.5	−1.5	0.2	− 0.7 = $g(BCD)$	BCD = −0.0875
+ + + +	2.3	−2.7	−1.2	0.3	0.1 = $g(ABCD)$	$ABCD$ = 0.0125
Total	57.5					
Sum of squares 219.15					3506.40/16=219.15	

The corresponding estimated effects are given by $g(\ldots)/(N/2)$. The algorithm is best explained with an example.

Example. The example shown in Table 47.24 has 2^4 runs. Thus the associated Yates algorithm will have six columns, as shown in Table 47.26. The grand average is $\bar{y} = 57.5/16 = 3.5938$. The next entry in column 6 is $g(A) = -12.9$. The estimated main effect of factor A is then

$$A \text{ effect} = \frac{-12.9}{8} = -1.6125$$

The estimate of the main effect of A can be checked by taking the average of the responses recorded on the high (+) side of the factor A and subtracting the average response on the low (−) side to give $\bar{y}_+ - \bar{y}_- = 22.3/8 - 35.2/8 = -12.9/8 = -1.6125$.

The remaining effects are similarly computed. Thus, the estimated AD interaction effect $= -2.5/8 = -0.3125$.

The following steps are checks on the computations in Table 47.26:

6. The sum of column 2 equals $g(T)$.

7. The sum of squares of the entries in column 2 equals 219.5. The sum of squares in column 6 divided by $2^4 = 3506.4/16 = 219.15$.

Testing Main Effects and Interactions. The grand average and 15 estimated effects obtained from the 2^4 design appear in the right-hand column of Table 47.26. The standard error (SE) of each estimated effect is given by

$$\text{SE(effect)} = \text{SE}(\bar{y}_+ - \bar{y}_-) = 2s\backslash\sqrt{N}$$

where N=total number of observations. The $100(1-\alpha)$ percent confidence limits are given by

$$\text{Effect} \pm t_{\alpha/2}\,[\text{SE(effect)}]$$

Needed is s^2, the estimate of the experimental error variance σ^2. An estimate of σ^2 can always be obtained from truly replicated trials, each set of replicates providing a single estimate of variance and then all the estimates pooled. However, in this example each trial was performed only once and some alternative procedure for securing an estimate of σ^2 is needed. We turn now to the analysis of variance table for the 2^k factorial design.

The Analysis of Variance for the 2^k Factorial. It is a rare design of experiments computer software program that omits the computations for a factorial design analysis of variance table. The computations are easy, though lengthy, when done by hand. We use our 2^4 example to demonstrate these computations for the interested reader. They are easily done with a computer spread sheet program.

The total variability of the observations is measured by $\Sigma(y_i - \bar{y})^2 = \Sigma y_i^2 - (\Sigma y_i)^2/N$, $i = 1, 2, \ldots, N$. In this example $\Sigma(y_i - \bar{y})^2 = 219.15 - (57.5)^2/16 = 12.51$. If this variability could be completely assignable to random errors, then the estimate of the variance σ^2 is given by $s^2 = \Sigma(y_i - \bar{y})^2/(N-1)$ with $(N-1)$ degrees of freedom. However, some of the movement amongst the observation y_i is likely caused by the influences of the controlled factors which make up the experimental design. The contribution of each factorial effect is given by its "sum of squares" $= N(\text{effect})^2/4$, labeled SSq, with one degree of freedom, as illustrated in Table 47.27.

The residual sum of squares represents variability remaining after all assignable causes have been subtracted. In Table 47.27 all 15 degrees of freedom with their associated SSq "sum of squares" are present, and thus the residual sum of squares and degrees of freedom are both zero.

To estimate the variance σ^2 in this unreplicated factorial design, we must declare some of the estimated effects to be manifestations of noise, i.e., not real effects and likely equal to zero. The most

TABLE 47.27 Analysis of Variance Table: 2^4 Factorial

Source of variation	SSq	DF
Total $= \Sigma(y_i - \bar{y})^2$	12.509375	15
A effect $= 16\,(-1.6125)^2/4$	10.400625	1
B effect $= 16\,(+0.3125)^2/4$	0.390625	1
C effect $= 16\,(-0.1125)^2/4$	0.050625	1
D effect $= 16\,(-0.1125)^2/4$	0.050625	1
AB effect $= 16\,(-0.4375)^2/4$	0.765625	1
AC effect $= 16\,(-0.0625)^2/4$	0.015625	1
AD effect $= 16\,(-0.3125)^2/4$	0.390625	1
BC effect $= 16\,(+0.1625)^2/4$	0.105625	1
BD effect $= 16\,(+0.0125)^2/4$	0.000625	1
CD effect $= 16\,(-0.0625)^2/4$	0.015625	1
ABC effect $= 16\,(+0.0625)^2/4$	0.015625	1
ABD effect $= 16\,(-0.2375)^2/4$	0.225625	1
ACD effect $= 16\,(-0.1125)^2/4$	0.050625	1
BCD effect $= 16\,(-0.0875)^2/4$	0.030625	1
$ABCD$ effect $= 16\,(+0.0125)^2/4$	0.000625	1
Residual sum of squares	0	0

reasonable collection of such effects is the three and four factor interactions. We thus sum their sum of squares and degrees of freedom and compose an estimate of variance $s^2 = 0.323125/5 = 0.064625$ with $\nu = 5$ degrees of freedom. The estimated standard deviation is $s = 0.254$. The standard error of an effect is then SE(effect) $= 2s/\sqrt{N} = 0.127$, and the $100(1 - \alpha)$ percent confidence limits for an effect are

$$\pm t_{\nu\alpha/2}\text{SE(effect)} = \pm\, t_{5,0.025}(0.127) = \pm 2.571(0.127) = \pm 0.326$$

Thus, an (approximate) 95 percent confidence interval for all remaining effects is given by effect ± 0.326. Any estimated effect whose confidence interval "effect ± 0.3265" includes zero may be declared not statistically significant. Of the estimated effects in this example only *A* and *AB* can be declared statistically significant and the effects *B* and *AD* nearly so. Computer software programs will perform analysis of variance *F* tests (or equivalent *t* tests) and compute their probabilities. These tests provide identical inferences. The analyst is advised to look at the interval statements rather than try to judge the differences between small probabilities. It is a rare individual who can really appreciate the true meaning of probabilities such as 0.07 versus 0.035.

A less formal analysis is to plot the absolute values of the estimated effects into a Pareto diagram. Here a Pareto diagram of the estimated effects indicates that factors *A, AB, AD,* and *B* are the "vital few." A Pareto diagram will identify "significant" effects almost as well as a collection of confidence intervals or a set of *F* tests in an ANOVA table (Hunter, W.G. 1977; Lenth 1989; Schmidt and Launsby 1991).

Collapsing the 2^k Factorials. Collapsing 2^k designs into lower "dimensionality" is an important strategy in the application of factorial designs. In practice it is very unlikely that all the possible main effects and interactions are real when the number of factors $k \geq 3$. In fact, particularly in screening situations where many factors are being co-studied, several individual factors may have no detectable main effect or interaction influences upon a response. Factorial designs for *k* modestly large are thus said to have "hidden replication," that is, an excess of degrees of freedom over and

TABLE 47.28 Collapsed 2^4 Giving a Replicated 2^3

A B D	Responses	Difference	
− − −	4.2, 3.9	0.3	
+ − −	3.1, 2.8	0.3	
− + −	4.5, 4.6	−0.1	
+ + −	2.9, 3.2	−0.3	$s^2 = \Sigma d^2/2n_d$
− − +	4.0, 4.0	0	
+ − +	3.0, 2.5	0.5	$s^2 = 0.57/2(8)$
− + +	5.0, 5.0	0	$= 0.03560$
+ + +	2.5, 2.3	0.2	

beyond those necessary to explain the response. Thus, when a factorial design is unreplicated, many estimated effects and their associated sums of squares and degrees of freedom may become available to employ in the estimate of variance. In this 2^4 example it appears that factor *C* makes no large contribution to the response, either as a main effect or as an interaction. The factor's effects are so small over its region of exploration as to be indistinguishable from noise. The 2^4 factorial now collapses into a replicated 2^3 design in the effective factors *A, B,* and *D.* The collapsed design is displayed in Table 47.28.

The estimated effects for factors *A, B,* and *D* remain unchanged. An estimate of variance can now be obtained from the repeated runs. When there are pairs of observations, a shortcut computation for s^2 is given by:

$$s^2 = \frac{\sum_i d_i^2}{2n_d}$$

where the d_i are the differences between the $i=1, 2, \ldots, d$ pairs of observations and n_d is the number of differences. This estimate of variance has $\nu=n_d$ degrees of freedom. Thus, for this example:

$$s^2 = \frac{0.57}{(2)(8)} = 0.0356$$

with $\nu = 8$ degrees of freedom and $s = 0.1887$. The standard error of each effect (excluding all those carrying the label *C,* of course), is

$$\text{SE(effect)} = \frac{2s}{\sqrt{N}} = \frac{2(0.1887)}{\sqrt{16}} = 0.0943$$

Using *t* with $\nu=8$ degrees of freedom, the 95 percent confidence limits for the estimated effects are given by

$$(t_{\alpha/2})\text{SE(effect)} = \pm 2.306(0.0943) = \pm 0.2175$$

Half-Normal Plots. Daniel (1959) proposed a simple and effective technique for use in the interpretation of data from the two-level factorial designs. This technique consists of plotting the absolute values of the estimated effects on normal probability paper (dropping the effect's signs), leading to "half-normal" plots. Effects indistinguishable from noise will fall along a straight line; effects that are statistically significant will fall well off the line. Full-normal plots consisting of the estimated effects *with* their signs are also possible; see Box, Hunter, and Hunter (1978). Normal probability plot routines are found in most statistical software programs. Half-normal plots are a particularly useful diagnostic tool offering evidence of wild observations and other conditions

that violate the assumptions of homogeneous variance, normality, and randomization. "Guardrails" can be used to judge the divergence of the plotted effects from the line; see Zahn (1975) and Taylor (1994). The Lilliefors (1967) test for normality can be helpful in judging all normal plots. For this and other nonparametric methods see Iman (1994).

EVOP: EVOLUTIONARY OPERATION

An important application of experimental design in the production environment was proposed by Box (1957). Essentially, a simple experimental design, run repeatedly, provides a routine of small systematic changes in a production process. The objective is to force the process to produce information about itself while simultaneously producing product to standards. Only small changes in the process factors are allowed, and the consequences of these changes must be detected in the presence of the many natural variabilities that surround the process. The repetition of an experimental design, commonly a 2^2 factorial with center point, permits the blocking of many of the disturbances that commonly influence production. Through the process of replication, the design provides steadily improving estimates of the main effects and interactions of the studied factors.

Response Surface. A response surface (see Box 1954) is a graphical representation of the connection between important independent variables, controlled factors, and a dependent variable. (An independent variable is a factor that is, or conceivably could be, controlled. Examples are flow rate and temperature. The value of a dependent variable is the result of the settings of one or more independent variables.) Most processes have several dependent variables, such as yield, impurities, and pounds per hour of a byproduct. These responses are usually smooth and may be graduated approximately by simple contours such as a family of lines or arcs. We ordinarily work on processes that have unknown response surfaces—if they were known, the work would not be necessary. See Response Surface Designs below.

A response surface for a process might look like the one in Figure 47.9, which shows the yield of a catalytic oxidation as a function of temperature and feed rate of hydrocarbon. If this information were known, the pounds per hour of product could be determined and better operating conditions selected for any desired production rate. The response surface is initially unknown, but improvement can be made if we only find out which way is up. Multiple regression can be used to approximate the response contours (see Section 44, under Multiple Regression).

EVOP Technique. The problem, then, is to increase profit in an operating plant with minimum work and risk and without upsetting the plant. These are the steps:

1. Survey company reports and open literature on the process. Study cost, yield, and production records.
2. Study this section on EVOP and preferably the definitive text (Box and Draper 1969).
3. Obtain agreement and support from production management. Organize a team and hold training sessions.
4. Select two or three controllable factors that are likely to influence the most important response.
5. Change these factors in repeated small steps according to a plan.
6. After the second repetition of the plan (Cycle 2) and each succeeding cycle, estimate the effects.
7. When one or more of the effects is significant, move to the indicated better operating conditions and start a new EVOP program, perhaps with new ranges or new factors.
8. After eight cycles, if no factor has been shown to be effective, change the ranges or select new variables (Box and Draper 1969).
9. Continue moving the midpoint of the EVOP plan and adjust the ranges as necessary.

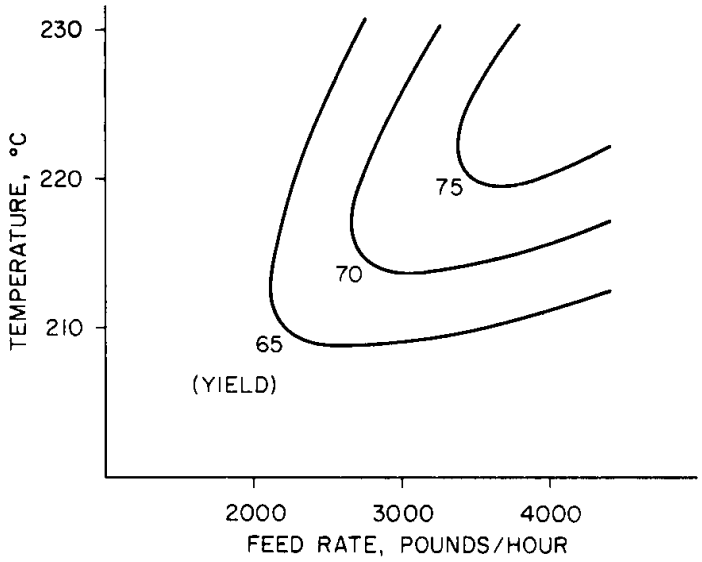


FIGURE 47.9 Typical response surface: yield as a function of feed rate and temperature for a catalytic reactor.

10. When a maximum has been obtained, or the rate of gain is too slow, drop the current factors from the plan and run a new plan with different factors.

The following topics explain these steps in detail.

Literature Search. Sources of information include the process instructions, company reports, manufacturers' literature, patents, textbooks, and encyclopedias of technology. Do not neglect people. Company personnel, consultants, and operators can all contribute. Search for information on

1. Important independent variables
2. Test methods for intermediate and final results
3. Recommended procedures
4. Records of good and bad results and their causes
5. Long-term history of results; plant production rate and yield by week or month and similar data; effect of past changes in equipment and conditions

When information is contradictory, an EVOP program is an ideal strategy for resolving the conflict. Always consider the physical and chemical principles that apply.

The EVOP Design. EVOP uses planned runs that are repeated over and over (replicated). One plan in wide use is the two-level complete factorial. There are important reasons to maintain observations on a known set of conditions, called a "reference point." For simplicity in the present discussion, let this point be the center of the square formed by the vertices of the factorial.

Example. This example shows coded data from an actual EVOP program (Barnett 1960). The problem involved a batch organic reaction, and after the steps above were followed, the two factors and their ranges were selected as shown in Figure 47.10. The response Y is a coded yield in pounds per batch and should be maximized. In this diagram, the reference run (batch) was made at 130°C for $3\frac{1}{2}$ h. The next batch was made at 120°C for 3 h, and so on. The first *cycle* contains five runs, one at each of the conditions. Samples were taken from each batch and analyses were obtained. If the process were continuous, it would be allowed to stabilize after each change of conditions.

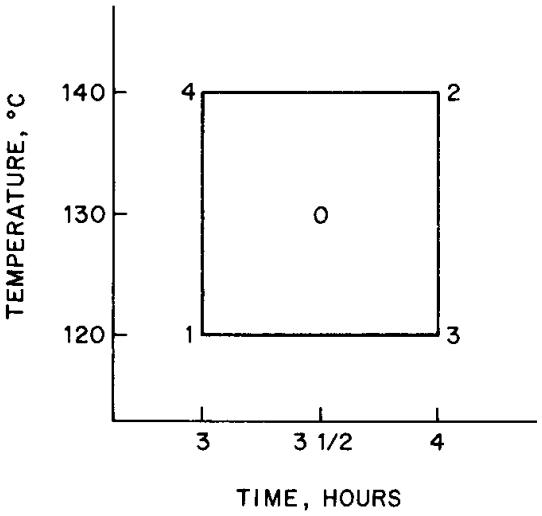


FIGURE 47.10 An EVOP plan. Numbers are in run order; 0 is the reference run.

Warning: It is not unusual to have difficulty in obtaining a representative sample. Procedures and tests are discussed in Sections 44 and 46.

Effects can be estimated at the end of Cycle 1, but in the absence of repeated trials, no estimate of error variance is available. At the conclusion of Cycle 2 both estimated effects and error variance are possible. The estimated effects and their confidence intervals may be computed on a form such as Figure 47.11, which shows this example. The form helps reduce the work and minimizes mistakes. Instructions should be printed on the reverse side (see Figure 47.12).

The error term that puts the magnitude of effects into perspective is obtained from the range by use of factor *K* first derived by Box and Hunter (1959). The uncertainty in estimating the effects is stated as a confidence interval and the Change-in-mean effect (CIM) is calculated by comparing the results at the outer four corners with the result in the center. The use of this last information is discussed below.

A "phase" is defined as all the cycles that use the same settings of the same factors. The phase average shows the general level of the response and can be used to compare different phases.

The left part of the form is used to record data and estimate effects. It also has the scaled diagram of the phase. The right third is used to determine the error limits of effects and corner aver-

J-227-x

CALCULATION OF AVERAGES

OPERATING CONDITIONS	0	1	2	3	4
SUM FROM PREVIOUS CYCLE	938	666	1104	904	948
AVERAGE FROM PREVIOUS CYCLE	938	666	1104	904	948
NEW OBSERVATIONS	792	840	883	825	996
DIFFERENCES (WATCH SIGNS)	+146	−174	+221	+79	−48
NEW SUMS (N.S.)	1730	1506	1987	1729	1944
NEW AVERAGES (N. S./N)	A) 865	B) 753	C) 994	D) 864	E) 972

CALCULATION OF EFFECTS

A EFFECT	B EFFECT	AB EFFECT
C 994 − B 753	C 994 − B 753	B 753 − D 864
D 864 − E 972	E 972 − D 864	C 994 − E 972
F 1858 − G 1725	1966 − 1617	1747 − 1836
1725 −	1617 −	− 1747
2 L 133 2 L	2 L 349 2 L	2 L 2 L 89
+ 66 −	+ 174 −	+ − 44

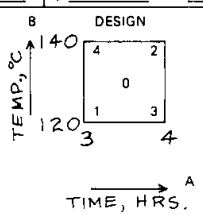
CHANGE-IN-MEAN EFFECT

F 1858 − A 865
G 1725 × 4
H 3583 − 3460
3460 −
5 L 123 5 L
+ 25 −

PHASE MEAN

H 3583
A 865
5 L 4448
890

DESIGN



REMARKS:

TEMPERATURE EFFECT IS BORDERLINE
RUN ANOTHER CYCLE

CALCULATION OF STANDARD DEVIATION

PREVIOUS SUM S = ——
NEW S = RANGE × K = 3.95 × 0.30 = 1.18
NEW SUM S = 1.18
NEW AVERAGE S_A = NEW SUM S/(N−1)
= 1.18 / 1 = 1.18
PREVIOUS AVERAGE S_A = ——

CALCULATION OF 2 S. E. ERROR LIMITS

FOR NEW AVERAGES AND NEW EFFECTS
L 1.41 × S_A 1.18 = ± 1.66

FOR CHANGE-IN-MEAN EFFECT
M 1.26 × S_A 1.18 = ± 1.49

FACTORS

N	K	L	M
2	0.30	1.41	1.26
3	0.35	1.15	1.03
4	0.37	1.00	0.89
5	0.38	0.89	0.80
6	0.39	0.82	0.73
7	0.40	0.76	0.68
8	0.40	0.71	0.63

CORNING GLASS WORKS
EVOLUTIONARY OPERATIONS
TWO VARIABLE WORK SHEET

PRODUCT EXAMPLE
RESPONSE YIELD, POUNDS
PHASE I CYCLE (N) 2
BY EHB DATE 1-11-71

MANUFACTURING AND ENGINEERING DIVISION

FIGURE 47.11 Calculations form at the end of Cycle 2. (*Modified from Barnett 1960.*)

Differences	Subtract new observations from old averages. Note the algebraic sign of the difference. (938 − 792 = 146)
New Sums	Add the new observations to the old sums. (938 + 792 = 1,730)
New Averages	Divide the new sums by *N*, the number of the cycle. (1730/2 = 865)
Calculation of Effects	(For example, the *A* effect) Write the new averages for operating conditions 2 and 3 opposite (C) and (D). Add these two to get the number in space (F). Carry out corresponding operations to get the number for space (G). The next operation is subtraction. If (F) is larger than (G), recopy (G) under (F) and subtract (G) from (F). If (G) is larger than (F), recopy (F) under (G) and subtract (F) from (G). In either case divide by 2 and the sign of quotient is shown on the form. (The A effect is +66.)
Change-in-mean effect	Copy (F) and (G) as shown and add these two. (A) is multiplied by 4. The next operation is subtraction as above. Divide by 5. (The Change-in-mean effect is +25.)
Average phase	Copy (H) and (A) from Change-in-mean box, add them, and divide by 5. (The phase mean is 890.)
Calculation of Standard Deviation	Range. The range is the algebraic difference between the most positive and most negative differences. The range is always positive. (The range of +146, −174, +221, +79, and −48 is 395.) The standard deviation is estimated by taking the range times *K*.
Constants	Read *K*, *L*, and *M* factors from the table.

Note: The numerical illustrations apply to the data of Fig. 47.11.

FIGURE 47.12 Instructions for EVOP form.

ages. The error limits are called "2 S.E." for "two standard errors." The estimated effect ±2 S.E. covers the usual 95 percent confidence region. Caution should be exercised in claiming statistical significance until after two or three cycles.

For the present example, at the end of Cycle 2 the *A* effect (time) is estimated as 66 ± 166; it lies somewhere between −100 and +232. Thus, the true value of this effect could be negative, positive, or nil. The *B* (temperature) effect, however, is estimated to be 174 ± 166, or in the range of +8 to 340. Technically, it is likely to be a positive real effect. The interaction *AB* is small, and so is the change in mean. Since the confidence region for *B* is so close to zero and following the advice above to be cautious at Cycle 2, another cycle is run. Its results are shown in Figure 47.13.

After Cycle 3, the *B* effect (temperature) was declared statistically significant, since its likely values fall in the range of 169 ± 92, or +77 to +261. It does not appear that data from more cycles would change the conclusion that temperature should be increased to increase the response *y*.

When statistical significance is found for one factor but not for the other, *move* the plan in the desirable direction for the "discovered" important factor and increase the range for the second (nonsignificant) one. Possible old and new phases of an EVOP program are shown in Figure 47.14. When significance is found for both variables, the center of the plan is moved in both directions in proportion to the size of the effects. This is the direction of steepest ascent (see Determine Direction of Steepest Ascent below, under First-Order Strategy). Whenever the plan is changed, a new *phase* is started. During the second and later phases, the previous estimate of standard deviation is often used since it was obtained under the current operating method.

Moves. To be conservative, as EVOP should be, moves are contiguous; i.e., one or more of the points in the old phase and new phase coincide. This limits the moves to those types shown in Figure 47.14.

There is nothing "magic" about drawing these plans as squares—1 h does not equal 20°C anyway. A particularly strong signal may justify a move to a plan that does not adjoin the previous one.

Change-in-Mean. The Change-in-mean effect is the difference between the results at the center point and the average of the other four peripheral points. It is therefore a signal of curvature as shown

J-227-x'

CALCULATION OF AVERAGES

OPERATING CONDITIONS	0	1	2	3	4
SUM FROM PREVIOUS CYCLE	1730	1506	1987	1729	1944
AVERAGE FROM PREVIOUS CYCLE	865	753	994	864	972
NEW OBSERVATIONS	934	853	1066	852	953
DIFFERENCES (WATCH SIGNS)	- 69	-100	- 72	+ 12	+ 19
NEW SUMS (N.S.)	2664	2359	3053	2581	2897
NEW AVERAGES (N. S./N)	A) 888	B) 786	C) 1018	D) 860	E) 966

CALCULATION OF EFFECTS

A EFFECT: C 1018, B 786, D 860, E 966, F 1878, G 1752, 1752, 126, 63

B EFFECT: C 1018, B 786, E 966, D 860, 1984, 1646, 338, 169

AB EFFECT: B 786, D 860, C 1018, E 966, 1804, 1826, 1804, 22, 11

CHANGE - IN - MEAN EFFECT: F 1878, A 888, G 1752, x 4, H 3630, 3552, 3630, 3552, 78, 16

PHASE MEAN: H 3630, A 888, 5 4518, 904

B DESIGN: TEMP, °C; 140 (4, 2, 0); 120 (1, 3); 3 4; A; TIME, HRS.

CALCULATION OF STANDARD DEVIATION

PREVIOUS SUM S = 118

NEW S = RANGE x K = 119 x 0.35 = 42

NEW SUM S - 160

NEW AVERAGE S A = NEW SUM S/(N-1) = 160 / 2 = 80

PREVIOUS AVERAGE S A = 118

CALCULATION OF 2 S. E. ERROR LIMITS

FOR NEW AVERAGES AND NEW EFFECTS: L 1.15 x S A 80 = ± 92

FOR CHANGE - IN - MEAN EFFECT: M 1.03 x S A 80 = ± 82

FACTORS

N	K	L	M
2	0.30	1.41	1.26
3	0.35	1.15	1.03
4	0.37	1.00	0.89
5	0.38	0.89	0.80
6	0.39	0.82	0.73
7	0.40	0.76	0.68
8	0.40	0.71	0.63

CORNING GLASS WORKS
EVOLUTIONARY OPERATIONS
TWO VARIABLE WORK SHEET

PRODUCT EXAMPLE
RESPONSE YIELD, POUNDS
PHASE I CYCLE (N) 3
BY EHB DATE 1-12-71

REMARKS:

TEMPERATURE EFFECT IS + 77 → + 261
INCREASE TEMPERATURE & START PHASE II

MANUFACTURING AND ENGINEERING DIVISION

FIGURE 47.13 Calculation form after Cycle 3.

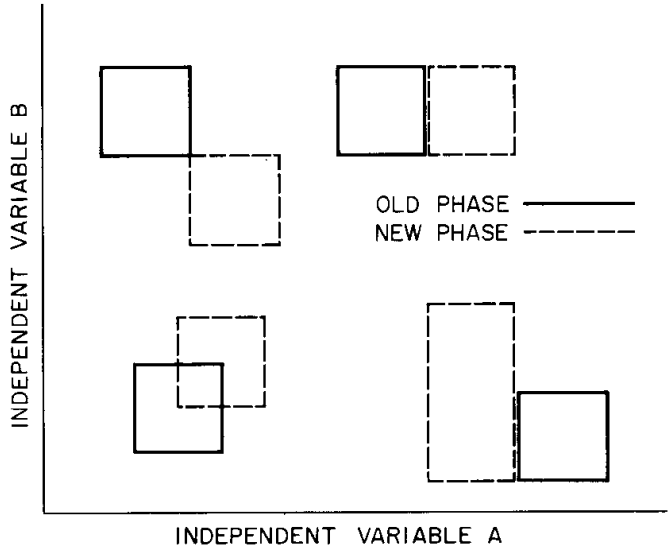


FIGURE 47.14 Possible relations of old phase and new phase.

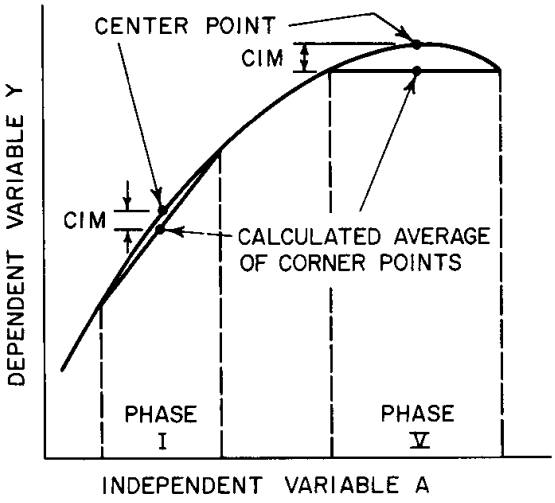


FIGURE 47.15 Cross section through a response surface. CIM indicates curvature.

in Figure 47.15. It is used in conjunction with the effects to indicate when a maximum (or a minimum) has been reached and to indicate the sensitivity of the response to changes in the independent factors. In the Taguchi literature the response would be termed "robust" to changes in the independent variable (Taguchi 1978). In rare cases it may happen that the first phase is located symmetrically about the maximum with respect to the two independent factors chosen. In this case the factors should be nonsignificant but the Change-in-mean may be significant.

Blocking. A process response ordinarily changes slightly with time, reflecting changes in sources of raw material, changes in air temperature from day to night, and so on.

Runs made close together in time are expected to be more nearly alike than those over a longer interval. Blocking is used to minimize the trouble caused by temporal changes of this type. For the EVOP calculations shown here, a block is one cycle. Changes in average level that occur between cycles are completely eliminated from the estimated effects, as can be seen by adding a constant to the five runs of Cycle 3, and recalculating the effects. (The phase average is changed, of course.)

Multiple Dependent Responses. So far the explanation has been in terms of a single dependent response. This is rather unrealistic, except for the profit variable. Most processes have several dependent responses that must be measured or calculated, such as yield, production rate, percent impurity, or pounds of byproduct. A calculation sheet is made for each dependent response, and statistical significance may be noted on one dependent sheet but not on others. In this case, it may be well to run another cycle or two to get more information on the other dependent responses before a move is made.

The most troublesome case occurs when the indicated directions for improvement of two responses (say production rate and percent impurity) do not agree. The EVOP program has brought information from the production process. Decisions as to what to do next now rest upon information supplied by the EVOP program coupled with information to be supplied by the subject matter experts.

BLOCKING THE 2^k FACTORIALS

Experimenters often find difficulty in maintaining a homogeneous experimental environment for all the experiments required in a 2^k factorial. For example, an experimenter might need 2 days to run the eight trials required in a 2^3 factorial. The question is how to choose the trials to be run each day so as not to disturb the estimates of the major effects of the three factors, i.e., how to "block" the design into two blocks of four runs each. Here blocking is accomplished by sacrificing the interaction estimate of least concern, i.e., the three-factor interaction. The procedure to be followed is illustrated in Table 47.29, part *a,* for a 2^3 factorial. First, the plus and minus signs of the 2^3 design are written down. Next, the columns of plus and minus signs commonly used to estimate the *ABC* interaction is constructed and labeled the block "generator." Those runs carrying a plus sign in the block generator column form the first block; those carrying a minus sign form the second block.

When this design is employed, the estimate of the three-factor interaction (abbreviated 3fi) cannot be distinguished from the block effect; the block effect and 3fi effect are "confounded." All other estimated effects are clear of the block effect.

TABLE 47.29 Partitioning the 2^3 Factorial Design

a. Partitioning the 2^3 into two blocks of four runs

2^3		2^3 in two blocks	
		(+) Block	(−) Block
A B C	*ABC* = block generator	*A B C*	*A B C*
− − −	−	+ − −	− − −
+ − −	+	− + −	+ + −
− + −	+	− − +	+ − +
+ + −	−	+ + +	− + +
− − +	+		
+ − +	−		
− + +	−		
+ + +	+		

b. Partitioning the 2^3 into four blocks of two runs

2^3		Block generators	
A B C		*AB*	*BC*
− − −		+	+
+ − −		−	+
− + −		−	−
+ + −		+	−
− − +		+	−
+ − +		−	−
− + +		−	+
+ + +		+	+
+ + Block	− + Block	+ − Block	− − Block
A B C	*A B C*	*A B C*	*A B C*
− − −	+ − −	+ + −	− + −
+ + +	− + +	− − +	+ − +

To partition the design into four blocks of two runs each, the proper block generators are provided by the two columns of plus and minus signs associated with the interactions *AB* and *BC* as illustrated in Table 47.29, part *b.* Note that the generators produce four combinations of minus and plus signs, each combination identifying a block of two runs. In this particular design all 2fi (two-factor interactions) are confounded with blocks.

The block generators must be carefully chosen. The blocking arrangements for the 2^3, 2^4, and 2^5 designs appear in Table 47.30. A more complete table and description of factorial design blocking appears in Box, Hunter, and Hunter (1978).

FRACTIONAL FACTORIAL EXPERIMENTS (EACH FACTOR AT TWO LEVELS)

If there are many factors, a complete factorial experiment, requiring all possible combinations of the levels of the factors, involves a large number of tests—even when only two levels of each fac-

TABLE 47.30 Blocking Arrangements for the 2^k Factorials

k = number of factors	Block size	Block generators	Interactions confounded with blocks
3	4	*ABC*	*ABC*
	2	*AB,BC*	*AB,BC, AC*
4	8	*ABCD*	*ABCD*
	4	*ABC, ACD*	*ABC, ACD,BD*
	2	*AB,BC,CD*	all 2fi and 4fi
5	16	*ABCDE*	*ABCDE*
	8	*ABC,CDE*	*ABC,CDE, ABDE*
	4	*ABC,BCD,CDE*	*ABC,BCD,CDE, AD, ABDE, BE, ACE*
	2	*AB,BC,CD,DE*	all 2fi and 4fi

tor are being investigated. In these cases, it is useful to have a plan that requires fewer tests than the complete factorial experiment. The fraction is a carefully prescribed subset of all possible combinations. The analysis of fractional factorials is relatively straightforward, and the use of a fractional factorial does not preclude the possibility of later completion of the full factorial experiment.

Confounding (Aliasing, Biasing). In a complete factorial experiment we have 2^k experimental trials. The 2^k experiments can be used to give independent estimates of all 2^k effects. In a fractional factorial (say the fraction $1/2^p$) there will be only 2^{k-p} experiments, and therefore only 2^{k-p} independent estimates are possible. In designing the fractional plans (i.e., in selecting an optimum subset of the 2^k total combinations), the goal is to keep each of the 2^{k-p} estimates as unbiased or "clear" as possible, i.e., to keep the estimates of main effects and if possible second-order interactions mutually unbiased, or nearly so.

To explain, consider the following 2^{3-1} fractional (the one-half, 2^{-1}, of the 2^3 factorial):

A B C	Observed
− − +	$y_1 = 8$
+ − −	$y_2 = 11$
− + −	$y_3 = 9$
+ + +	$y_4 = 14$

The main effects are given by the statistics $\bar{y}_+ - \bar{y}_-$, where once again the plus and minus subscripts of each letter in the design identify the observations entering each average. Thus, the main effect of *A* is estimated to be $(11 + 14)/2 - (8 + 9)/2 = 4.0$. The main effects of *B* and *C* are, respectively, $(9 + 14)/2 - (8 + 11)/2 = 2$ and $(8 + 14)/2 - (11 + 9)/2 = 1.0$. Now consider the estimate of the two-factor interaction *AB*. The analyst will find that the signs required to estimate the *AB* interaction are identical to those already employed to estimate the main effect of *C*. The main effect of *C* and the two-factor interaction *AB* are *confounded.* Said another way, the statistic $\bar{y}_+ - \bar{y}_- = 1.0$ has an "alias" structure; that is, the statistic may be identified as either *C* or *AB*. In fact, the expected value of the statistic equals *C*+*AB,* the sum of the two effects, and in the absence of clear information on the main effect of *C,* we cannot tell whether the *AB* effect is plus, minus, large, or small. The reader will note that estimate *A* is confounded with *BC,* as is *B* with *AC*.

When some or all main effects are confounded with two-factor interactions, the fractional factorial design is said to be of "Resolution III." When one or more of the main effects are confounded

with (at least) three-factor interactions, the fractional is said to be a "Resolution IV" design. Fractionals with main effects confounded with (at least) four-factor interactions are of "Resolution V," etc. (See Box and Hunter 1959).

DESIGNING A FRACTIONAL FACTORIAL DESIGN

Let N equal the number of runs and k the number of factors to be investigated. When $N = 2^k$, we have a full factorial design. When $N = 2^{k-p}$, we have a $(1/2)^p$ replicate of the 2^k factorial; for example, a 2^{7-3} is a one-eighth replicate of a 2^7 factorial and contains 16 runs.

To design a one-half replicate design in N runs, first write down (Yates order is best) the full factorial design in $(k - 1)$ factors. Next write down the column of signs associated with the highest-order interaction. These signs are now used to define the versions of the kth factor. For example, to construct the 2^{4-1} design, begin with a 2^3 factorial in factors *A, B,* and *C* as illustrated in Table 47.31. Next to the columns for *A, B,* and *C* write down the column of signs associated with the *ABC* interaction. Use these signs to identify the two versions of factor *D.* (The other one-half fraction is obtained by reversing the signs of the column *ABC.*)

TABLE 47.31 Constructing the 2^{4-1} Fractional Factorial*

Generator		Principal design	Alternative design†
A B C	*ABC = D*	*A B C D*	*A B C D*
− − −	−	− − − −	− − − +
+ − −	+	+ − − +	+ − − −
− + −	+	− + − +	− + − −
+ + −	−	+ + − −	+ + − +
− − +	+	− − + +	− − + −
+ − +	−	+ − + −	+ − + +
− + +	−	− + + −	− + + +
+ + +	+	+ + + +	+ + + −

*Example run: Run no. 2 requires the experimenter to hold factor *A* at +, factor *B* at −, factor *C* at −, and factor *D* at +.

†The alternative fraction is obtained by reversing the signs of the *ABC* vector, that is, by setting $D = -ABC$.

To construct a one-quarter replicate design, two columns of signs are required in addition to the standard factorial in N runs; the one-eighth replicate design requires three additional columns, etc. The columns of signs to be used must be carefully chosen; they are listed in Table 47.32 for designs up to $k = 7$ factors. Table 47.32 is an adaptation of a much more extensive table given in Box, Hunter, and Hunter (1978). Extensive listings of fractional factorial designs can also be found in Diamond (1989).

Most design of experiment software programs provide two-level fractional factorial designs of any desired resolution along with the alias-confounding patterns associated with each estimated effect. Given below is an example of the construction of a 2^{4-1} and design and analysis of a 2^{6-2} factorial.

Example. To construct a fractional factorial design for $k = 6$ factors in $N = 16$ runs, first write down the full factorial 2^4 design in factors *A, B, C,* and *D.* Consulting Table 47.32, the vectors of plus and minus signs associated with the interaction *ABC* are now used to define the versions of factor *E.* The signs of the *BCD* interaction are similarly used to define the versions of factor *F.* The completed 2^{6-2} design is displayed in Table 47.33 along with observed responses, Yates' algorithm, and identified effects.

TABLE 47.32 Vectors Used for the Construction of Fractionals

Number of runs N	Number of factors k				
	3	4	5	6	7
4	2^{3-1} $\pm AB = C$	NA	NA	NA	NA
8		2^{4-1} $\pm ABC = D$	2^{5-2} $\pm AB = D$ $\pm AC = E$	2^{6-3} $\pm AB = D$ $\pm AC = E$ $\pm BC = F$	2^{7-4} $\pm AB = D$ $\pm AC = E$ $\pm BC = F$ $\pm ABC = G$
16			2^{5-1} $\pm ABCD = E$	2^{6-2} $\pm ABC = E$ $\pm BCD = F$	2^{7-3} $\pm ABC = E$ $\pm BCD = F$ $\pm ACD = G$
32				2^{6-1} $\pm ABCDE = F$	2^{7-2} $\pm ABCD = F$ $\pm ABDE = G$
64					2^{7-1} $\pm ABCDEF = G$

TABLE 47.33 A 2^{6-2} Resolution IV Fractional Factorial and Associated Yates Analysis

Generators: $E=ABC$ and $F=BCD$
Defining relation: $I=ABCE=BCDF=ADEF$

$A\ B\ C\ D\ E\ F$	Obs.*		Yates algorithm			Effects	Identification†
− − − − − −	124	271	541	1137	2405	150.3125	Average
+ − − − + −	147	270	596	1268	11	1.375	$A + BCE + ABCDF + DEF$
− + − − + +	145	284	615	−1	35	4.375	$B + ACE + CDF + ABDEF$
+ + − − − +	125	312	653	12	−139	−17.375	$AB + CE + ACDF + BDEF$
− − + − + +	138	307	3	27	93	11.625	$C + ABE + BDF + ACDEF$
+ − + − − +	146	308	−4	8	−1	−0.125	$AC + BE + ABDF + CDEF$
− + + − − −	162	323	3	−63	35	4.375	$BC + AE + DF + ABCDEF$
+ + + − + −	150	330	9	−76	169	21.125	$ABC + E + ADF + BCDEF$
− − − + − +	125	23	−1	55	131	16.375	$D + ABCDE + BCF + AEF$
+ − − + + +	182	−20	28	38	13	1.625	$AD + BCDE + ABCF + EF$
− + − + + −	181	8	1	−7	−19	−2.375	$BD + ACDE + CF + ABEF$
+ + − + − −	127	−12	7	6	−13	−1.625	$ABD + CDE + ACF + BEF$
− − + + + −	168	57	−43	29	−17	−2.125	$CD + ABDE + BF + ACEF$
+ − + + − −	155	−54	−20	6	13	1.625	$ACD + BDE + ABF + CEF$
− + + + − +	154	−13	−111	23	−23	−2.875	$BCD + ADE + F + ABCEF$
+ + + + + +	176	22	35	146	123	15.375	$ABCD + DE + AF + BCEF$

*Obs.=observations.

†Expected value of the effect from the defining relation.

Identifying the Estimates. The 2^{6-2} design was generated by setting $E = ABC$ and $F = BCD$. A simple procedure for identifying the biases (aliases) of the effects estimable from this design is as follows. Multiply the expression $E=ABC$ by E and the expression $F=BCD$ by F. This gives $E^2 = ABCE$ and $F^2 = BCDF$. Now adopt the rule that whenever a symbol appears squared, it is replaced by an I, the "identity," a symbol equivalent to the numeral 1. We now have for the design "generators" $I = ABCE$ and $I = BCDF$. Multiplying the generators together gives the *defining relation* $I = ABCE = BCDF = AB^2C^2DEF$, which reduces to $I = ABCE = BCDF = ADEF$.

When Yates' algorithm is applied to the 16 runs of the 2^{6-2}, the algorithm estimates 15 effects and provides each with its initial name, as illustrated in Table 47.33. The defining relation is now employed to determine the additional names (aliases or biases) of each of these statistics. Thus, the statistic labeled the "main effect" of A actually equals $A = BCE = ABCDF = DEF$, an expression obtained by multiplying through the defining relation by the symbol A. Similarly, the statistic initially called the "ABC interaction" actually estimates $ABC = E = ADF = BCDEF$. The estimates and their full names are given in Table 47.33.

Five of the estimates appear unusually large and are good candidates for measured phenomena distinguishable from natural variability (noise). Using only their first- and second-order names we have: -17.375 estimates $AB + CE$, 11.625 estimates C, 21.125 estimates E, 16.375 estimates D, and 15.375 estimates $DE + AF$. A reasonable interpretation of these statistics is that factors C, D, and E have detectable important influences upon the response over their studied ranges, while factors A and B do not. This conclusion obviously needs confirmation, but it represents a good first guess. The 2^{6-2} design now collapses into a 2^3 factorial repeated in factors C, D, and E.

OTHER FRACTIONAL FACTORIALS

Although the 2^{k-p} fractional factorial designs discussed here are the most frequently used, many other fractionals exist. For example the Plackett and Burman (1966) designs are two-level fractional factorials whose number of runs N is not a power of 2 but a multiple of 4. The $N = 12$ design for $k \leq 11$ factor design is displayed in Table 47.34. The templates for producing the designs for $N = 20$, 24, 28, and 36 can be found in Box and Draper (1987) and Myers and Montgomery (1995).

All Plackett and Burman designs are Resolution III. The alias structure associating main effects and two factors interactions is not as readily available as those of the regular $2_{\text{III}}^{k-\text{p}}$ designs. However, *every* Resolution III design can be made into a Resolution IV design by the principle of "fold-over."

TABLE 47.34 A Plackett and Burman Design

($k = 11$ factors, $N = 12$ runs)

A	B	C	D	E	F	G	H	I	J	K
+	−	+	−	−	−	+	+	+	−	+
+	+	−	+	−	−	−	+	+	+	−
−	+	+	−	+	−	−	−	+	+	+
+	−	+	+	−	+	−	−	−	+	+
+	+	−	+	+	−	+	−	−	−	+
+	+	+	−	+	+	−	+	−	−	−
−	+	+	+	−	+	+	−	+	−	−
−	−	+	+	+	−	+	+	−	+	−
−	−	−	+	+	+	−	+	+	−	+
+	−	−	−	+	+	+	−	+	+	−
−	+	−	−	−	+	+	+	−	+	+
−	−	−	−	−	−	−	−	−	−	−

To "fold over" a design one merely writes it down again with all signs reversed. A fold-over design combined together with its original Resolution III design forms a design of Resolution IV.

Fractional factorials are not limited to the 2^{k-p} designs. The 3^3 factorial design can be reduced into a variety of fractions including the Latin Square, see Hunter (1985*a*) for graphical displays. Mixed level fractional factorials are also available; see Addelman (1962). Many novel fractionals have been published: Hahn and Shapiro (1966), Webb (1968*a, b*), Margolin (1969), Anderson and Thomas (1978), and Rechtschaffner (1967). Fractional factorials when the number of runs is less than the number of variables, "supersaturated" designs, are also possible; see Booth and Cox (1962) and Lin (1993, 1995).

Several design of experiments computer software programs will construct unique fractional factorial designs, allowing the experimenter to not only choose the number of runs, but also the number of factors and levels. The reader can only be warned that the application of designs with small numbers of runs and many factors assumes great simplicity in the mathematical model for the response function under study, most particularly that only main effects exist and that all interactions are either zero or truly near zero. One is also well advised to leave some redundancy in one's design, i.e., extra runs to provide degrees of freedom that can be employed in estimating the experimental error variance σ^2 [see Snee (1985) and Berk and Picard (1991)]. (There has never been a signal in the absence of noise, and one should plan on measuring the noise as well as possible signals.

Screening Experiments. The saturated 2^{k-p}_{III} designs for studying k factors in $N = k + 1$ runs can be of value in early screening efforts to detect important factors among many candidates. The possible biasing influences of interactions is very serious in these applications and should always be kept in mind. Certain software programs can construct fractional factorial designs to provide estimates of all k main effects and certain prechosen two-factor interactions. However, the assumption that all interactions of importance can be announced prior to the experimental program being designed and run can be naive when one considers the number of such interactions that may be possible. Furthermore, the discovery of interactions can easily be as important as the identification of main effects. Conservative practice in the use of fractional factorials generally requires designs of Resolution IV, i.e., the ability to separate main effects from two-factor interactions. See Hurley (1994); Tang and Tang (1994); and Haaland and O'Connell (1995).

Orthogonal Arrays. The terminology "orthogonal array," used by the earliest creators of balanced block experimental designs, has been popularized by Taguchi (1987); see below. All the 2^k, 2^{k-p}, Plackett and Burman, and Latin square type designs can be called "orthogonal arrays." The number of runs associated with each orthogonal array is often identified by the notation $L(N)$, as for example the $L8$ orthogonal array is the 2^3 (or 2^{7-4}) two-level factorial and the $L9$ and the $L27$ are the 3^2 and 3^3 factorials. The $L36$ can be viewed either as a three-level design or as a 6×6 Latin square.

The classification "orthogonal array" is appropriate to any experimental design that can provide estimates of effects having zero correlations. The designs are sometimes described as "main effect clear" designs, although they are often adapted to take into account certain interactions. The 2^{k-p} and 3^{k-p} fractionals, the Latin square designs (and the Graeco-Latin and Hyper-Graeco-Latin square designs) thus qualify, as do the mixed-level factorials and fractional factorials and the balanced block designs. The Box and Behnken (1960) designs form novel fractions of the three-level orthogonal arrays. The terminology "orthogonal arrays" recognizes the geometric multidimensional nature of all these designs; that is, in the N-dimensional space of the observations, the vectors representing the effects to be estimated are all mutually perpendicular. One orthogonal array design popularized by Taguchi is the $L9$ design, the 3^{4-2} for studying four factors each at three levels in nine runs. A critique of the application of this and other three-level orthogonal arrays is found in Hunter (1985).

In listing orthogonal array designs the Taguchi literature uses the notation (1, 2) and (1, 2, 3) to identify the levels (versions) of each variable instead of the geometric notation $(-1, +1)$ and $(-1, 0, +1)$. Examples of the analysis of orthogonal arrays employing the Taguchi terminology and methodology are provided by Barker (1990, 1994), Kacker (1985), Phadke et al. (1983), Phadke (1989), and Taguchi (1978, 1987). These authors pay particular attention to the use of "inner" and

"outer" orthogonal arrays, or in a parallel terminology, to "design" and "noise" matrices. The designs are similar in structure to the classical split-plot designs.

To construct unique designs for estimating main effects and *selected* interactions, Taguchi employs "linear graphs" associated with each orthogonal array; see Kacker and Tsui (1990), Wu and Chen (1991), and Wu, Mao, and Ma (1990). Linear graphs provide a geometric analogue to the use of fractional factorial defining relations (see Identifying the Estimates under Designing a Fractional Factorial Design, above). Almost always, classical methods employing defining relations can provide experimental designs identical to or better than those provided by the application of linear graphs.

TAGUCHI OFF-LINE QUALITY CONTROL

The Japanese engineer–quality expert Genechi Taguchi must be credited with much of today's interest in the use of factorial and fractional factorial designs on the part of the automotive, communication, and assembly industries; see Taguchi (1978). Within these industrial environments experiments are run to identify the settings of both product design parameters and process variables that will simultaneously provide a manufactured item whose response is robust to process variability while meeting the customer's product expectations and possible environmental challenges. The adaptation of statistical experimental design to these objectives has its origins in Taguchi's early work in the communications industries in Japan in the 1950s. The strategy is called "parameter design" or "robust design." It is important to note that the word "design" takes differing connotations: product design, process design, and statistical design.

Taguchi requires manufactured products be created to meet the following criteria:

1. To protect the product from sources of variability occurring within the manufacturing process
2. To have minimum variability about the customer's target values
3. To be robust to environmental factors encountered by the customer

More formally, a product's response y is considered to be a function of "controllable" factors x and "noise" factors z. The objective is to choose settings of x that will make the product's response y insensitive to variability associated with both x and z and still meet target specifications with least variability.

Inner and Outer Arrays. The statistical designs associated with the Taguchi approach to product and process design usually contain both an "inner" and "outer" array, or "design matrix" and "noise matrix," each constructed from the orthogonal arrays. (See previous section.) The inner array consists of a statistical experimental design employing the controllable factors x, while the outer array is a statistical experimental design in the noise factors z which are now intentionally varied. (Occasional x factors may also be included as "noise" factors in an outer array.) The entire design forms a split-plot-like experimental array with the z outer array repeated within each of the settings of the x inner array. For example, if the inner array were a $L16 = 2^{8-4}$ and the outer array a $L9 = 3^{4-2}$ there would be a total of $16 \times 9 = 144$ experiments, each of the 16 runs of the $L16$ containing its own 9-run $L9$ design. Experimental designs providing for inner and outer arrays that employ fractional factorial arrangements are also possible; see Shoemaker, Tsui, and Wu (1991) and Montgomery (1991).

At each setting of the inner array Taguchi now determines a "signal to noise" statistic composed from the outer array, noise matrix, observations. Taguchi focuses on a quadratic loss function $Q = (\eta - \theta)^2 + \sigma^2$, where η is the expected product response, θ the target value (the quantity $\eta - \theta$ is bias), and σ^2 the variance of the observed responses. To aid in minimizing the loss function, Taguchi defines the "signal to noise" ratio SN, where commonly $\text{SN} = 10 \log_{10} (\eta/\sigma)^2$, $\eta = E(y)$, and $\sigma^2 = \text{Var}\ (y)$. At each setting of the inner array the statistic $\text{SN} = 10 \log_{10} (\bar{y}/s)^2$ is computed using the n observations from the outer array occurring only at that setting. Other definitions for

SN are also suggested. For example, when a higher response is preferred, Taguchi proposes SN=10 $\log_{10}[(y_1^2 + y_2^2 + \cdots + y_n^2)/n]$ and, for lower desired response, the statistic SN = 10 $\log_{10}[(1/y_1^2 + 1/y_2^2 + \cdots + 1/y_n^2)/n]$, where $y_1, y_2, \ldots, y_n$ are the n observations from the outer array unique to each setting of the inner array.

Most statisticians recommend that the averages and estimated variances obtained at each of the points of the inner array be separately analyzed. Members of the Taguchi school continue to recommend the analysis of the various signal to noise statistics. No closure to the debate seems imminent. One thing is clear. The fraternity of quality engineers and statisticians is indebted to Prof. Taguchi for proposing the concept of the design of robust products (parameter design) and for adapting the arts of statistical design of experiments to that end use.

A large body of literature exists describing and offering examples of the Taguchi approach. An excellent summary and critique of the methodology identified with Prof. Taguchi, and possessing an extensive bibliography, appears in a discussion organized by Nair (1992). Major authors identified with the Taguchi approach are Taguchi (1978, 1986, 1987); Kacker (1985); Kacker and Shoemaker (1986); Kacker and Tsui (1990); Barker (1990, 1994); Phadke (1989); Phadke et al. (1983); Leon, Shoemaker, and Kacker (1987); and their various coworkers. Authors who have discussed the Taguchi work include Bisgaard (1993*b*), Box (1988), Box and Jones (1986), Box and Myers (1986), Easterling (1985), Goh (1993), Grove and Davis (1991*b*), Hunter (1985*a,* 1989), Hurley (1994), Lucas (1985), Miller et al. (1993), Montgomery (1991), Nair and Shoemaker (1990), Stephens (1994), Tribus and Sconyi (1989), and Vining and Myers (1990).

RESPONSE SURFACE DESIGNS

Response Surface Methodology (RSM) has been successfully used to optimize many different kinds of industrial units, processes, and systems. It is an experimental approach and has been applied in research and development laboratories and sometimes on actual plant equipment itself. In the latter situation, however, Evolutionary Operation is often more appropriate. Evolutionary Operation is an alternative form of RSM that is useful for both objectives of screening and optimizing.

RSM experimental designs require that important factors influencing a process, identified perhaps by a screening experiment, be varied in a carefully chosen pattern of experiments. Commonly two controlled variables and a single response variable are studied. The data obtained are then analyzed with the primary objective of providing a rough map (usually a contour representation) of the response surface over the region of the controlled variables investigated. The mathematical models employed are the first-order and second-order polynomials. Thus, the fitted response surface may be planar (a first-order approximation to the "true" surface) or nonplanar or curved (a second-order approximation). The fitted models are obtained using ordinary least squares estimation procedures (regression analysis). Often a fitted response surface will suggest alternative levels of a factor to provide better yields. Thus, a program of RSM may go through several stages of mapping before "best" conditions are identified. When more than two controlled variables are studied, contour surfaces are employed. See Box and Draper (1987). Nor is it necessary to map only a single response. Two or more responses can be separately mapped and their maps superimposed to identify regions of "optimum" operability. Finding "optimum" operating conditions does not always mean finding the factor settings that give the biggest or smallest response. Suggestions for further reading include Box, Hunter, and Hunter (1978); Box and Draper (1987); Khuri and Cornell (1987); Mason, Gunst, and Hess (1989); Haaland (1989); and Myers and Montgomery (1995). A history of RSM appears in Myers, Khuri, and Carter (1989).

Modern computers and software programs have made RSM a most useful and valuable statistical tool. Not only are the burdens of computation minimized, but the ability of computers to display maps of the fitted response surfaces provides the analyst with vivid insights into the nature of the responses and factors under investigation. Most software programs will allow an experimenter to obtain a first- or second-order mapping of an unknown response surface employing almost any collection of data; all that is needed is a good least-squares regression program. However, good experimentation

requires a careful selection of points. A poorly designed response surface program is analogous to viewing a scene through an astigmatic lens. The consequence is a warped view. In most circumstances a first- or second-order rotatable (nonastigmatic) response surface design offers the best strategy. Fitting response surfaces to haphazardly acquired data is called PARC analysis by some statisticians (PARC: Practical Accumulated Records Computations, or Planning After Research is Completed.) The use of the standard first- and second-order RSM designs for $k = 2$ factors and a single response is described here.

In implementing RSM, a number of statistical procedures discussed in other sections of this Handbook are used. The concept of RSM was first developed and described by Box and Wilson (1951). At first RSM was used primarily as an experimental optimization technique in the chemical industry. Since then, however, it has found application in many other fields (see Hill and Hunter 1966).

RSM can be usefully regarded as consisting of two stages:

1. First-order stage, in which a first-order mathematical model is contemplated, a factorial or other first-order design performed, the data fitted, the contours of the response drawn, and the direction of steepest ascent determined and pursued
2. Second-order stage, in which a second-order mathematical model is contemplated, a central-composite or other second-order design performed, the data fitted, the contours drawn, a canonical analysis performed, and an optimum located

Response Surface Methodology is actually more flexible than these brief definitions indicate. A skeletal outline, which shows some of the possible paths through an RSM study, is given in flow-diagram form in Figure 47.16.

Weakness of One-Variable-at-a-Time Approach. A popular method of experimentation is the one-factor-at-a-time approach. Each factor, in turn, is varied while all the rest of the factors are held at some fixed, constant levels. One trouble with this approach is that a false optimum can be reached. Consider the following hypothetical illustration.

Example. Under study is a chemical reaction in which there are two factors of interest, the concentration of one of the reactants and the reaction time. What settings for these two factors will maximize the yield? The best known settings, at the outset of the investigation, are a concentration of 25 percent and a time of 1 h (see Figure 47.17).

Following a one-factor-at-a-time approach, the engineer first runs a series of experiments by varying the time, while holding the concentration at 25 percent. The results show that a maximum yield of about 65 percent is obtained when the time is 1.9 h (position *E* on the line *A* to *B* in Figure 47.17). Holding the time fixed at this value, varying concentration along the line *C* to *D,* and obtaining a maximum at 25 percent, the engineer reaches the conclusion that the maximum yield (65 percent) is achieved when the concentration is 25 percent and the time is 1.9 h. This conclusion, however, is incorrect.

Response Surface Approach. The actual situation, unknown to the experimenter, is shown in Figure 47.17. Here the yield is shown as a function of both concentration and time. The solid curved lines in the figure are contour lines of constant yield. For example, there is an entire set of conditions of concentration and time that give an 80 percent yield. The contour surface can be viewed as a mountain; the peak of the mountain is the point *P.* The contours of 90, 80, and so forth, can be viewed as altitudes. These numbers represent the percentage yields.

The engineer's objective was to find those settings for the concentration and time that would give the maximum yield. Viewed geometrically, what the engineer was trying to do was to climb to the highest point on the mountain. The attempt failed for a fairly simple reason.

Figuratively speaking, by varying time, the engineer first traversed the hill going along a path from point *A* to point *B* (see Figure 47.17). Between *A* and *E* the path led up the mountain, but then

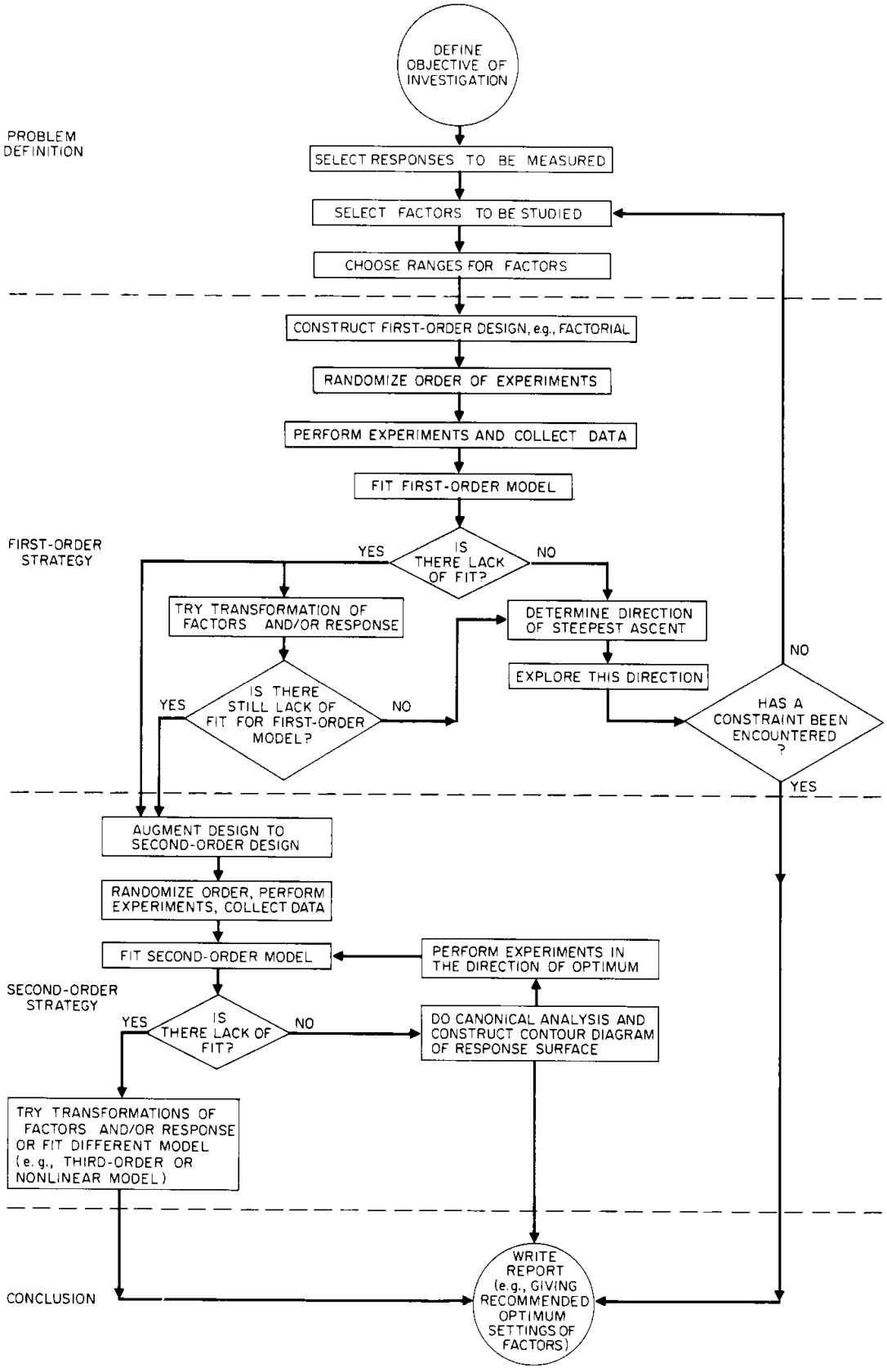


FIGURE 47.16 Outline of main ideas of Response Surface Methodology.

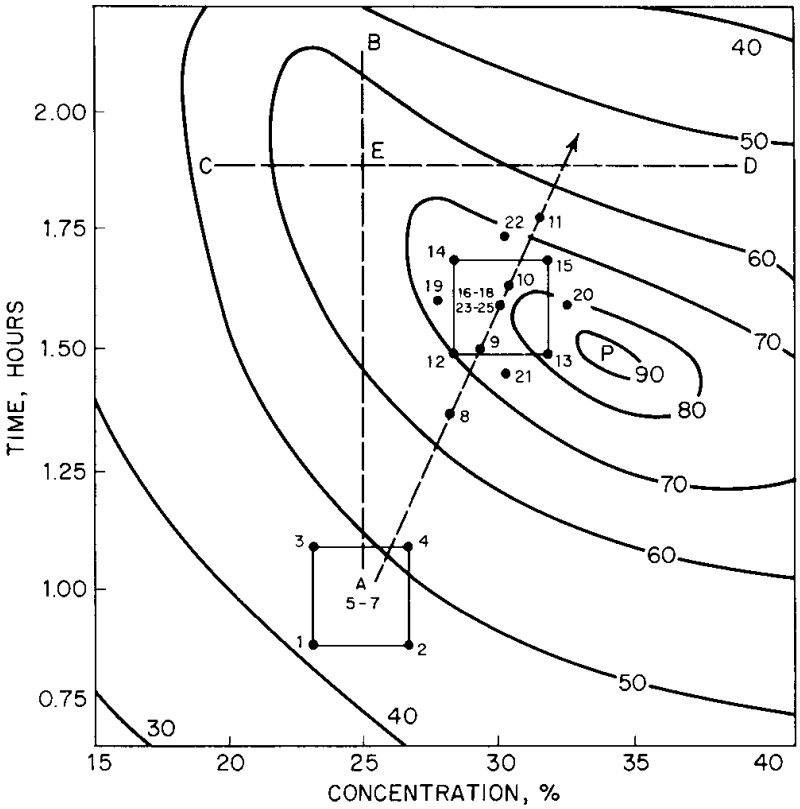


FIGURE 47.17 Response surface showing yield of a chemical reaction as a function of concentration and time.

at point *E* it started to go down the other side. From point *E* to point *B* the engineer was walking down the other side of the hill. The traverse for varying concentration (*C* to *E* to *D*) is shown.

The experimenter has achieved a yield of only 65 percent (at *E*), whereas a yield in excess of 90 percent (at *P*) is possible. This higher yield can be achieved by *simultaneously* increasing concentration and decreasing time from the experimenter's reported "optimum" values.

If the contours of the hill were circular and there were no experimental error, this one-at-a-time-procedure would have taken the engineer to the highest point on the hill. In general, the contours of real response surfaces are not circular nor experimental error (noise) absent, and thus what is needed is a more sophisticated experimental strategy such as RSM.

Beginning of Program. The RSM approach (see Figure 47.16) will now be applied to the example of maximizing the yield.

Define Objective of Investigation. It is of the utmost importance to define clearly the objective of the study to be undertaken. It is surprising how often in practice this step is either ignored or not given the careful attention it deserves. This often leads to difficulties later on. In the present example the objective is to maximize the yield. The objective, in general, may involve multiple criteria, that is, to maximize yield while simultaneously meeting other objectives such as minimizing impurity and obtaining an acceptable range of viscosity.

Select Factors and Ranges. The next step is to select the factors to be studied together with the ranges over which they are to be studied. It is necessary to understand the technical aspects of the experimental situation for this to be done intelligently. The specific *scale* over which each factor is to be studied must also be chosen. For example, instead of varying time linearly in units of hours, the experimenter might choose the basic scale to be the logarithm of the number of

hours. In the present example, the variables concentration and time are selected. Initially, it is decided to vary concentration from 23 to 27 percent and time from 0.9 to 1.1.

First-Order Strategy

Construct Design and Collect Data. The 2^2 factorial design with three center points, shown in Tables 47.35, is constructed. [Further discussion of the number of center points and other matters on setting up the design is given in Cochran and Cox (1957), Hunter (1959), Box and Draper (1987), and Myers and Montgomery (1995).] The order of the seven runs is randomized, the experiments are performed, and the results shown in Table 47.35 are obtained. The results are displayed in Figure 47.18.

Fit First-Order Model and Check for Lack of Fit. The analysis of these results can be carried out in either one of two equivalent ways. The effects and interaction of the factorial design can be calculated with their associated 95 percent confidence intervals, as is also shown in Table 47.35.

TABLE 47.35 Results of First-Order Design

Run number	X_1 = concentration		X_2 = time		Y = yield
	%	Coded units	Hours	Coded units	%
1	23	−1	0.9	−1	43.7
2	27	+1	0.9	−1	44.5
3	23	−1	1.1	+1	47.2
4	27	+1	1.1	+1	51.8
5	25	0	1.0	0	46.8
6	25	0	1.0	0	45.9
7	25	0	1.0	0	45.3

Calculation of main effects

Concentration:	$(-Y_1 + Y_2 - Y_3 + Y_4)/2 = (-43.7 + 44.5 - 47.2 + 51.8)/2 = 2.7$
Time:	$(-Y_1 - Y_2 + Y_3 + Y_4)/2 = (-43.7 - 44.5 + 47.2 + 51.8)/2 = 5.4$
Interaction:	$(+Y_1 - Y_2 - Y_3 + Y_4)/2 = (43.7 - 44.5 - 47.2 + 51.8)/2 = 1.9$
Curvature:	$(+Y_1 + Y_2 + Y_3 + Y_4)/4 - (Y_5 + Y_6 + Y_7)/3 = 46.8 - 46.0 = 0.8$

Calculation of confidence intervals

Concentration:	$\pm 2ts/\sqrt{n} = \pm 2(4.30)(0.755)/\sqrt{4} = \pm 3.25$
Time:	$\pm 2ts/\sqrt{n} = \pm 2(4.30)(0.755)/\sqrt{4} = \pm 3.25$
Interaction:	$\pm 2ts/\sqrt{n} = \pm 2(4.30)(0.755)/\sqrt{4} = \pm 3.25$
Curvature:	$\pm ts\sqrt{(1/n) + (1/n_0)} \equiv \pm(4.30)(0.755)\sqrt{(¼) + (⅓)} = \pm 2.48$

Effects	Calculated 95% confidence interval, % yield
Concentration	2.7 ± 3.25
Time	5.4 ± 3.25
Interaction	1.9 ± 3.25
Curvature	0.8 ± 2.48

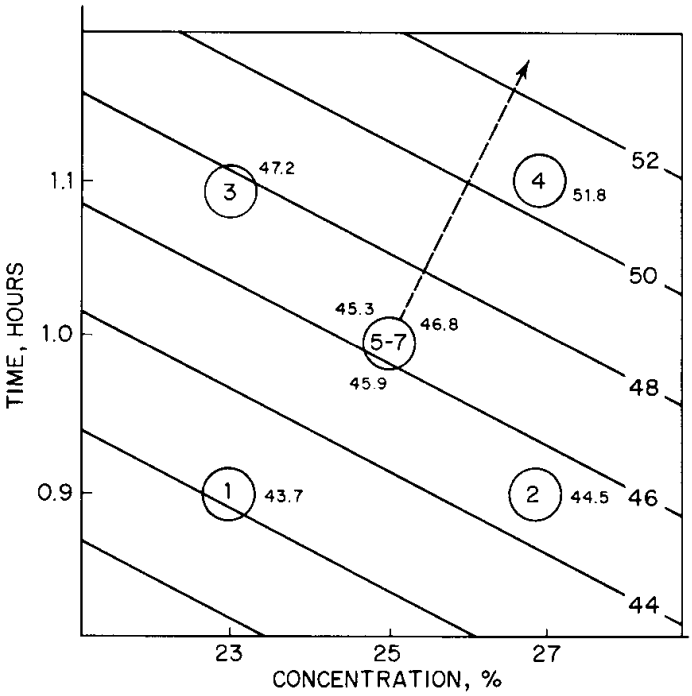


FIGURE 47.18 Results of first-order design with fitted first-order (planar) response surface.

As the center-point conditions have been repeated three times, an estimate of the variance can be readily obtained. (If repeat runs have not been performed, it might be possible to obtain an appropriate estimate in some other way, for example, from some external source, past experience, or from a technique like half-normal plots; see Daniel 1959.) With the three values 46.8, 45.9, and 45.3, $s^2 = 0.57$ is calculated as an estimate of the variance of an individual observation with two degrees of freedom.

Employing the estimated standard deviation $s = 0.755$ with two degrees of freedom, the confidence intervals suggest that the main effects of concentration and time are significant while the estimates of second-order effects (interaction and curvature) are indistinguishable from zero. We are thus able to employ as an approximation to the unknown response surface the first-order model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

where Y is the observed response and the β's are coefficients to be estimated from the data. The quantities X_1 and X_2 are independent variables representing the experimental factors concentration and time, where

$$X_1 = \frac{\text{concentration}(\%) - 25}{2}$$

$$X_2 = \frac{\text{hours} - 1.0}{0.1}$$

These expressions for X_1 and X_2 code the original settings of concentration and time to match those of the 2^2 factorial design with center point given in Table 47.37 and displayed in Figure 47.18. For example, when concentration = 23 and time = 0.9 hours, $X_1 = -1$ and $X_2 = -1$. The quantity ϵ in the model is assumed to be a random error normally distributed, independent, with constant variance σ^2. Although standard regression techniques (see Section 44) can be used to fit the

first-order model, estimates of the β's are readily obtained as a consequence of the 2^2 factorial design with center point. The estimate of β_0 is is the average, the estimate of β_1 one half the concentration effect, and the estimate of β_2 one half the time effect. Thus the fitted equation becomes

$$\hat{Y} = 46.46 + 1.35X_1 + 2.70X_2$$

The equation in terms of the original variables is

$$\hat{Y} = 28.23 + 0.675[\text{conc\%}] + 1.35[\text{hours}]$$

Setting $\hat{Y} = 44.0$ will produce the straight line contour labelled 44 in Figure 47.18. The fitted contours suggest the response surface to be well represented by the plane.

$$\hat{Y} = B_0 + B_1X_1 + B_2X_2 = 46.46 + 1.35X_1 + 2.70X_2$$

in the X_1, X_2 coordinate system.

A second method for evaluating the fit is to use the analysis of variance. The resulting ANOVA table (Table 47.36) indicates that the first-order model (above equation) adequately fits the data. (See Section 44; also Draper and Smith 1981.) The ratio of the lack of fit mean square divided by the pure error mean square is 4.13, and since this value is less than $F_{2.2}(0.95) = 19.0$, there is no evidence of lack of fit of the first-order model. Since there is no evident lack of fit, it is reasonable to study the implications of the fitted first-order model (above equation). The plane described by this equation is represented in Figure 47.18 by the straight contour lines.

Determine Direction of Steepest Ascent. The direction of steepest ascent is indicated in Figure 47.18. (For further details on direction of steepest ascent, see Cochran and Cox 1957, p. 357, and Box and Draper 1987.) It is perpendicular to the contour lines. Four experiments (numbers 8 to 11) in this direction indicate that the center of a second design should be approximately at a concentration of 31 percent and a time of 1.6 h. The design employed and the data obtained after performing the runs in random order are shown in Table 47.37 as runs 12 to 18. An analysis of the data shows apparent lack of fit (Table 47.38). The ratio of the lack of fit mean square divided by the pure error mean square is 26.8, and since this value is greater than $F_{2.2}(0.95) = 19.0$, there is evidence of lack of fit of the first-order model.

Second-Order Strategy

Construct Design and Collect Data. Since lack of fit is detected, the design is augmented by adding runs 19 to 25 to form the second-order (central composite) design shown in Table 47.37. In general,

TABLE 47.36 ANOVA Table: First-Order Model, First Design*

Source	Sum of squares	Degrees of freedom	Mean square
Mean b_0	15,107.86	1	
b_1	7.29	1	7.29
b_2	29.16	1	29.16
Lack of fit	4.71	2	2.36
Pure error	1.14	2	0.57
Total	15,150.16	7	

*In the literature of response surface methodology, it is customary that the ANOVA table include a term for the sum of squares for the mean. In other uses of ANOVA, some authors and computer software programs exclude the sum of squares for the mean.

TABLE 47.37 Results of Second-Order Design

Run number	Concentration (coded units) X_1	Time (coded units) X_2	Yield Y, %
	First-order design		
12	−1	−1	69.3
13	+1	−1	85.1
14	−1	+1	72.8
15	+1	+1	73.6
16	0	0	80.9
17	0	0	78.4
18	0	0	80.4
	Augmenting runs		
19	$-\sqrt{2}$	0	71.4
20	$+\sqrt{2}$	0	78.9
21	0	$-\sqrt{2}$	73.9
22	0	$+\sqrt{2}$	69.1
23	0	0	76.4
24	0	0	78.5
25	0	0	76.3

TABLE 47.38 ANOVA Table: First-Order Model, Second Design

Source	Sum of squares	Degrees of freedom	Mean square
Mean b_0	41,734.32	1	
b_1	68.89	1	68.89
b_2	16.00	1	16.00
Lack of fit	94.12	2	47.06
Pure error	3.50	2	1.75
Total	41,916.83	7	

if a model does not fit, it may be advantageous, instead of immediately considering a higher-order model, to consider transformations of the factors and/or the responses. See Box and Cox (1964), Box and Tidwell (1962), and Draper and Hunter (1967).

Fit Second-Order Model and Check for Lack of Fit. The fitted second-order equation obtained by least squares is

$$\hat{Y} = 78.50 + 3.40X_1 - 1.85X_2 - 3.75X_1X_2 - 1.21X_1^2 - 3.03X_2^2$$

The contours of this equation are shown in Figure 47.19 with the second-order design results. No lack of fit is evident from either visual inspection or statistical calculation (see Table 47.39). The form of the above equation can be simplified so the shape of the response surface can be better appreciated. It is difficult to visualize the surface from the equation because it contains six constants. A canonical analysis, which involves a translation and rotation of the coordinates from the original (X_1, X_2) axes to the new (Z_1, Z_2) axes, gives an equation containing only three constants:

$$Y - 173.83 = -0.0332Z_1^2 - 8.4075Z_2^2$$

This equation indicates that because of the negative coefficients for Z_1^2 and Z_2^2, the fitted response surface has a maximum point. A direction in which to proceed at the next stage to search for the maximum is indicated by the arrow in Figure 47.19. The arrow points toward the "top of the mountain." The investigation might terminate after experimenting in this direction, perhaps with a few added points in the vicinity of the maximum. In some situations it may be useful to perform a full

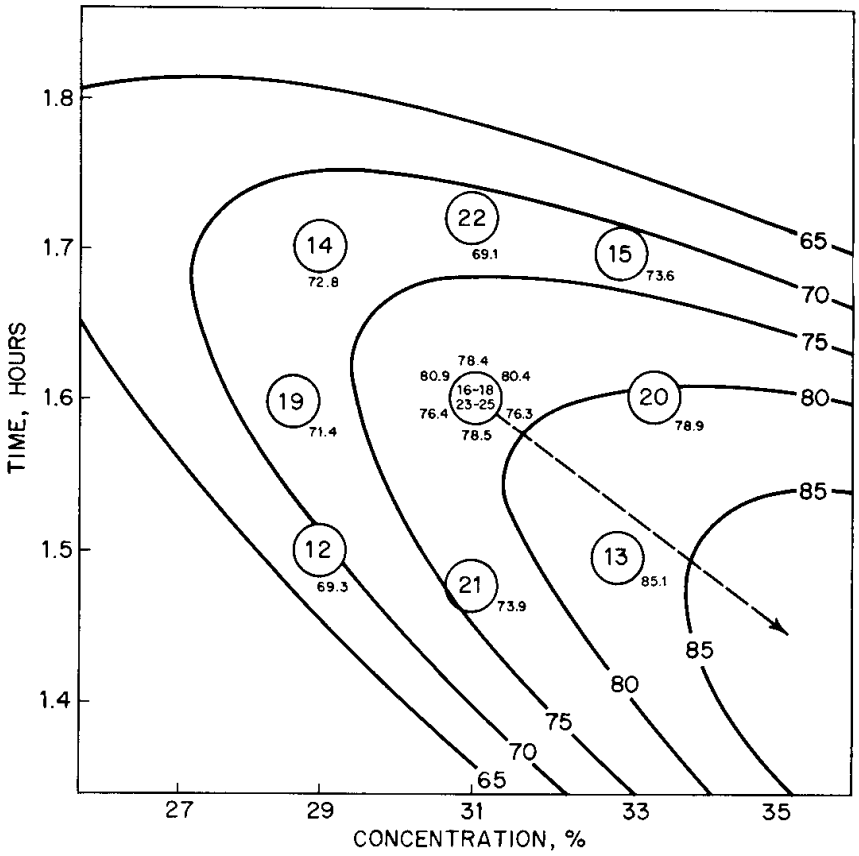


FIGURE 47.19 Results of second-order design with fitted second-order (nonplanar) response surface.

TABLE 47.39 ANOVA Table: Second-Order Model

Source	Sum of squares	Degrees of freedom	Mean square
Mean b_0	81,016.07	1	
First-order b_i	119.87	2	59.94
Pure second-order b_{ii}	80.73	2	40.37
Mixed second-order b_{ij}	56.25	1	56.25
Lack of fit	2.23	3	0.74
Pure error	21.77	5	4.35
Total	81,296.92	14	

Note: The ratio of the lack of fit mean square divided by the pure error mean square is 0.17, and since this value is less than $F_{0.95}$ for 3, 5 degrees of freedom (5.41), there is no evidence of lack of fit of the second-order model.

second-order design near the final optimum. [For further details on canonical analysis and RSM in general, see Box and Draper (1987) and Myers and Montgomery (1995).]

Many response surface experimental designs are available, in particular the three-level factorials, the central composite, the rotatable designs, and Box-Behnken designs. For response surface designs using a minimum number of runs while preserving many of the qualities of the larger designs see Draper (1985) and Draper and Lin (1990). Some computer software programs can provide unique response surface designs (commonly *D*-optimal) constructed to match special constraints provided by the experimenter.

MIXTURE DESIGNS

In some experiments with mixtures, the property of interest depends on the proportions of the mixture components and not on the amounts (volume or weight) of the individual components. For example, stainless steel is a mixture of different metals, and its tensile strength depends on the proportions of the metallic elements present; gasoline is ordinarily a blend of various stocks, and the octane rating of the final blend depends on the proportions going into the blend. The proportions of the components of a mixture must add up to unity, and in the most general case the proportion of any component may range from zero to unity. An important reference text is Cornell (1990).

In the design of mixtures, the factor space available for experimentation is constrained, since the proportions used must sum to unity. It has been shown that if the number of components in the mixture is *q,* the available factor space becomes a regular $(q - 1)$-dimensional simplex (e.g., a triangle for $q = 3$), a tetrahedron for $q = 4$).

A natural approach would be to take a uniformly spaced distribution of experimental points over the available factor space. This results in the simplex lattice designs proposed by Scheffé (1958). A (q, m) lattice, for example, is a lattice for q components, where the proportions for each component have $m+1$ equally spaced values from 0 to 1, i.e., the values 0, $1/m$, $2/m$, etc. For three components, the proportions of each component would be 0, $1/2$, 1 when $m=2$; and 0, $1/2$, $2/3$, 1 when $m = 3$. The lattice resulting when $m = 2$ is called the quadratic lattice, the lattice resulting when $m = 3$ is called the cubic lattice, etc. (see Figure 47.20).

In addition, modified lattices can be made by adding center points to the two-dimensional face or faces of the quadratic lattice. This provides a useful design called the "special cubic lattice."

The number of points k required for any lattice except the special cubic is found by using the formula

$$k = \frac{(m + q-1)!}{m!(q - 1)!}$$

The number of points required for the special cubic is

$$k = \frac{q(q + 1)}{2} + \frac{q(q - 1)(q - 2)}{6}$$

The number of points required for several values of m and q is given in Table 47.40.

The property of interest is measured at each of the design points (corresponding to mixtures of different proportions). Simplified polynomials are used to relate the response variable y to the various mixture proportions used.

Another useful design called the "special cubic" by Scheffé (1958) requires seven points for three-component mixtures—the six points of a $(q = 3, m = 2)$ lattice plus a seventh point at $X_1 = 1/3$, $X_2 = 1/3$, $X_3 = 1/3$.

The seven mixtures are the three pure components, the three binary mixtures, and the ternary mixture, as shown in Table 47.41.

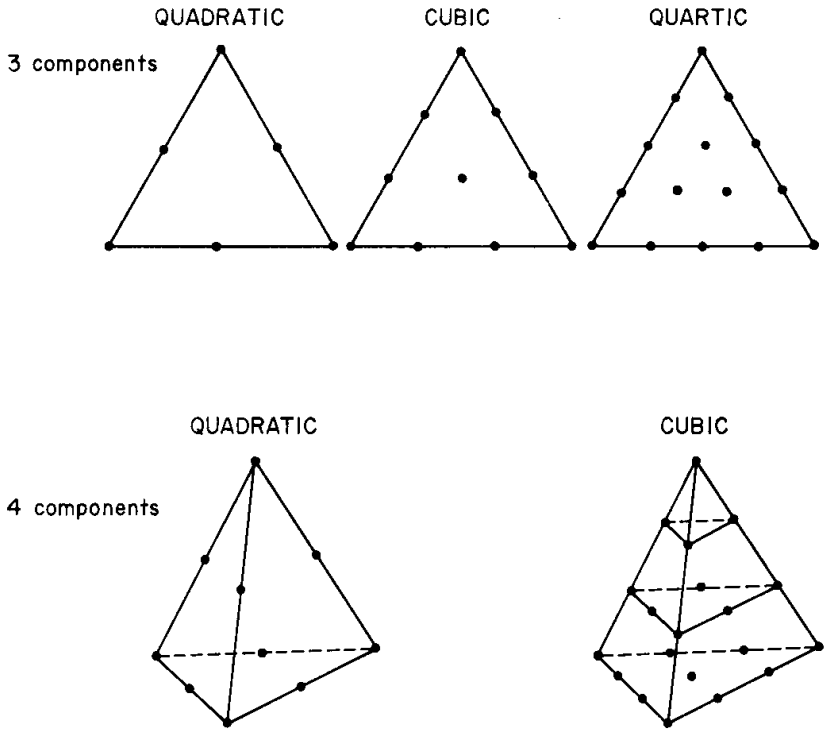


FIGURE 47.20 Lattice designs for three- and four-component mixtures. (*Reprinted with permission from Gorman and Hinman 1962.*)

TABLE 47.40 Number of Points Required for Lattice Designs

Number of components q	Type of lattice			
	Quadratic, $m = 2$	Special cubic, $m = 2$	Cubic, $m = 3$	Quartic, $m = 4$
3	6	7	10	15
4	10	14	20	35
5	15	25	35	70
6	21	41	56	126
8	36	92	120	330

TABLE 47.41 Design Points for Special Cubic (Three-Component Mixture)

Point number	x_1	x_2	x_3	Response
1	1	0	0	y_1
2	0	1	0	y_2
3	0	0	1	y_3
4	½	½	0	y_4
5	½	0	½	y_5
6	0	½	½	y_6
7	⅓	⅓	⅓	y_7

The "special cubic" corresponds to the equation:

$$y = B_1X_1 + B_2X_2 + B_3X_3 + B_{12}X_1X_2 + B_{13}X_1X_3 + B_{23}X_2X_3 + B_{123}X_1X_2X_3$$

The computed coefficients are

$$b_1 = y_1 \qquad b_2 = y_2 \qquad b_3 = y_3$$

$$b_{12} = 4y_4 - 2(y_1 + y_2)$$

$$b_{13} = 4y_5 - 2(y_1 + y_3)$$

$$b_{23} = 4y_6 - 2(y_2 + y_3)$$

$$b_{123} = 27y_7 - 12(y_4 + y_5 + y_6) + 3(y_1 + y_2 + y_3)$$

The subject of mixture designs now has a vast body of literature, and many computer software programs devote attention to both their design and analysis. Computer printouts of contour representations are particularly valuable. Factors composing mixture designs are often constrained to fall within narrow ranges, thus forming isolated mixture regions and requiring novel experimental designs difficult to obtain without a computer. Constrained factors are sometimes combined with factors not constrained, once again leading to unique designs and analyses. The designs may also be blocked and run sequentially. The textbook by Cornell (1990*a*) is devoted entirely to the topic of mixture experiments. See also Myers and Montgomery (1995). Important papers are Gorman and Hinman (1962), Thompson and Myers (1968), Snee (1973, 1979), Crosier (1984), Piepel and Cornell (1994), and Draper et al. (1993).

GROUP SCREENING DESIGNS

Novel experimental designs for finding the few effective factors out of a very large number of possible factors have been called "group screening designs." These designs have the following structure: groups are formed, each containing several factors; the groups are tested; and individual factors of the groups that prove to contain significant factors are then separately tested. Such designs, proposed by Connor (1961) and further studied by Watson (1961), are intended to minimize the amount of experimentation required.

The experimental variables are divided into groups, and each group is treated as a single variable until an effect on the response variable is shown.

The following assumptions are made:

1. All factors initially have the same probability of being effective.
2. The factors do not interact.
3. The directions of effects, if they exist, are known.

The number of factors is $f = gk$, where g = number of groups and k = number of factors per group. For example, consider an experiment with nine factors, which are divided into three groups of three factors each (i.e., $g = 3$, $k = 3$). The upper and lower levels of the groups are defined as follows:

1. Group factor *X* consists of factors *A, B, C.*

   Level 1: All three factors at lower level (0, 0, 0)

   Level x: All three factors at upper level (1, 1, 1)

2. Group factor *Y* consists of factors *D, E, F.*

   Level 1: All three factors at lower level (0, 0, 0)

Level y: All three factors at upper level (1, 1, 1)

3. Group factor *Z* consists of factors *G, H, I.*

Level 1: All factors at lower level (0, 0, 0)

Level z: All factors at upper level (1, 1, 1)

The first-stage design studies the *group* factors, for example by using a half-replicate of a 2^3 factorial. This requires the four group treatment combinations *x, y, z,* and *xyz* corresponding to treatment combinations for the nine factors as follows:

x(1,1,1 0,0,0,0,0,0)

y(0,0,0 1,1,1 0,0,0)

z(0,0,0 0,0,0 1,1,1)

xyz(1,1,1 1,1,1 1,1,1)

The results of the first-stage experiment will indicate which group factors contain at least one effective factor. A second-stage experiment, which may consist of a half-replicate of a 2^3, will then be run on each effective group factor to determine which of the individual factors are effective. For further details, see Watson (1961). Patel (1962) gives detailed procedures for two-, three-, and four-stage screening tests.

The application of group screening designs that has been discussed here is to the identification of effective experimental factors, but there is extensive literature relating to the screening of effective responses, e.g., to compounds and drugs, and to the group testing of individuals. Papers of interest are by Ehrenfeld (1972), Pocock (1983), Mundel (1984), Hwang (1984), and Hayre (1985). An excellent review of the entire problem of group screening is provided by Tang and Tang (1994).

REFERENCES

Addelman, S. (1962). "Symmetrical and asymmetrical fractional factorial plans." *Technometrics,* vol. 4, pp. 47–58.

American Society for Quality, Statistics Division, ASQ Quality Press, Milwaukee.

1: *How To Analyze Data with Simple Plots,* Wayne Nelson
2: *How To Perform Continuous Sampling,* 2nd ed., K. S. Stephens
3: *How To Test for Normality and Other Distribution Assumptions,* S. S. Shapiro
4: *How To Perform Skip-lot and Chain Sampling,* 2nd ed., K. S. Stephens
5: *How To Run Mixture Experiments for Product Quality,* J. A. Cornell
6: *How To Analyze Reliability Data,* W. Nelson
7: *How To and When to Perform Bayesian Acceptance Sampling,* T. W. Calvin
8: *How To Apply Response Surface Methods,* J. A. Cornell
9: *How To Use Regression Analysis in Quality Control,* D. C. Crocker
10: *How To Plan an Accelerated Life Test—Some Practical Guidelines,* W. Meeker and G. Hahn
11: *How To Perform Statistical Tolerance Analysis,* N. Cox
12: *How To Choose a Proper Sample Size,* G. G. Brush
13: *How To Use Sequential Statistical Methods,* T. P. McWilliams
14: *How To Construct Fractional Factorial Experiments,* R. F. Gunst and R. L. Mason
15: *How To Determine Sample Size and Estimate Failure Rate in Life Testing,* E. C. Moura
16: *How To Detect and Handle Outliers,* B. Iglewicz and D. C. Hoaglin

Anderson, D. A., and Thomas, A. M. (1978). "Resolution IV fractional factorial designs for the general asymmetric factorial." *Communications in Statistics,* vol. A8, pp. 931–943.

Automotive Industry Action Group (1990). *Measurement System Analysis Reference Manual.* Chrysler, General Motors, and Ford, Detroit.

Asao, M. (1992). "A simulation game of experimental design." *Quality Engineering,* vol. 4, pp. 497–517.

Bainbridge, T. R. (1965). "Staggered, nested designs for estimating variance components." *Industrial Quality Control,* vol. 22, pp. 12–20.

Barker, T. B. (1990). *Engineering Quality by Design: Interpreting the Taguchi Approach.* Marcel Dekker, New York.

Barker, T. B. (1994). *Quality by Experimental Design,* 2nd ed., Marcel Dekker, New York.

Barnett, E. H. (1960). "Introduction to Evolutionary Operation." *Industrial Engineering Chemistry,* vol. 52, pp. 500–503.

Barrentine, L. B. (1991). *Concepts for R&R Studies.* ASQC Quality Press, Milwaukee.

Bechhofer, R. E., and Tamhane, H. E. (1983). "Design of experiments for comparing treatments with a control: Tables of optimal allocation of observations." *Technometrics,* vol. 25, pp. 87–95.

Beckman, R. J., and Cook, R. D. (1983). "Outlier…s." *Technometrics,* vol. 25, pp. 119–149.

Berk, K. N., and Picard, R. R. (1991). "Significance tests for saturated orthogonal arrays." *Journal of Quality Technology,* vol. 23, pp. 73–89.

Bicking, C. A. (1954). "Some uses of statistics in the planning of experiments." *Industrial Quality Control,* vol. 10, pp. 23–31.

Bisgaard, S. (1992). "Industrial use of statistically designed experiments: Case study references and some historical anecdotes." *Quality Engineering,* vol. 4, pp. 547-562.

Bisgaard, S. (1993*a*). "Spreadsheets for the analysis of two-level factorials." *Quality Engineering,* vol. 6, pp. 149–157.

Bisgaard, S. (1993*b*). "Iterative analysis of data from two-level factorials." *Quality Engineering,* vol. 6, pp. 319–330.

Bishop, T., Peterson, B., and Trayser, D. (1982). "Another look at the statistician's role in experimental planning and design." *The American Statistician,* vol. 36, pp. 387–389.

Booth, K. H. V., and Cox, D. R. (1962). "Some systematic supersaturated experimental designs." *Technometrics,* vol. 4, pp. 489–495.

Bowman, K. O., Hopp, T. H., Kacker, R. N., and Lundegard, R. J. (1991). "Statistical quality control techniques in Japan." *Chance,* vol. 4, pp. 15–21.

Box, G. E. P. (1954). "The exploration and exploitation of response surfaces: some general considerations and examples." *Biometrics,* vol. 10, pp. 16–36.

Box, G. E. P. (1957). "Evolutionary Operation: A method for increasing industrial productivity." *Applied Statistics,* vol. 6, pp. 81–101.

Box, G. E. P. (1976). "Science and Statistics." *Journal of the American Statistical Association,* vol. 71, pp. 791–799.

Box, G. E. P. (1988). "Signal to noise ratios, performance criteria and transformations" (with discussion). *Technometrics,* vol. 30, pp. 1–40.

Box, G. E. P. (1990). "Do interactions matter?" *Quality Engineering,* vol. 2, pp. 365–369.

Box, G. E. P. (1991). "Finding bad values in factorial designs." *Quality Engineering,* vol. 3, pp. 405–410.

Box, G. E. P. (1992). "Sequential experimentation and sequential assembly of designs." *Quality Engineering,* vol. 5, pp. 321–330.

Box, G. E. P. (1992). "What can we find out from eight experimental runs?" *Quality Engineering,* vol. 4, pp. 619–627.

Box, G. E. P. (1992). "What can we find out from sixteen experimental runs?" *Quality Engineering,* vol. 5, pp. 167–181.

Box, G. E. P., and Behnken, D. W. (1960). "Some three level designs for the study of quantitative variables." *Technometrics,* vol. 2, pp. 477–482.

Box, G. E. P., Bisgaard, S., and Fung, C. A. (1988). "An explanation and critique of Taguchi's contributions to quality engineering." *Quality and Reliability Engineering International,* vol. 4, pp. 123–131.

Box, G. E. P., and Cox, D. R. (1964). "The analysis of transformations." *Journal of the Royal Statistical Society, Series B,* vol. 26, p. 211.

Box, G. E. P., and Draper, N. R. (1969). *Evolutionary Operation.* John Wiley, New York.

Box, G. E. P., and Draper, N. R. (1987). *Empirical Model Building and Response Surfaces.* John Wiley, New York.

Box, G. E. P., and Hunter, J. S. (1959). "Condensed calculations for evolutionary operation programs". *Technometrics,* vol. 1, pp. 77–95.

Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978). *Statistics for Experimenters,* John Wiley, New York.

Box, G. E. P., and Jones, S. (1986). "Split plot designs for robust product experimentation." *Journal of Applied Statistics,* vol. 19, pp. 2–36.

Box, G. E. P., and Myers, R. D. (1986). "An analysis for unreplicated fractional factorials." *Technometrics,* vol. 28, pp. 11–18.

Box, G. E. P., and Myers, R. D. (1986). "Dispersion effects from fractional designs." *Technometrics,* vol. 28, pp. 19–27.

Box, G. E. P., and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Reading, MA.

Box, G. E. P., and Tidwell, P. W. (1962). "Transformations of the independent variable." *Technometrics,* vol. 4, p. 531.

Box, G. E. P., and Wilson, K. B. (1951). "On the experimental attainment of optimum conditions." *Journal of the Royal Statistical Society, Series B,* vol. 13, p. 1.

Burdick, R. K., and Graybill, F. A. (1992). *Confidence Intervals on Variance Components,* Marcel Dekker, New York.

Chambers, J. M., and Hastie, T. J. (1992). *Statistical methods in S,* Wadsworth, Pacific Grove, CA.

Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. (1993). *Graphical Methods for Data Analysis,* Chapman and Hall, New York.

Clatworthy, W. H. (1973). *Tables of the Two-Associate-Class Partially Balanced Designs.* National Bureau of Standards Applied Mathematics Services Publication 63, U.S. Government Printing Office, Washington, DC.

Cleveland, W. S. (1993). *Visualizing Data,* Hobart Press, Summmit, NJ.

Cochran, W. G., and Cox, G. M. (1957). *Experimental Designs,* 2nd ed., John Wiley, New York.

Coleman, D. E., and Montgomery, D. C. (1993). "A systematic approach to planning for a designed industrial experiment." *Technometrics,* vol. 35 (with discussion), pp. 1–27.

Connor, W. S. (1961). "Group screening designs." *Industrial and Engineering Chemistry,* vol. 53, pp. 69A–70A.

Conover, W. J. (1980). *Practical Non-Parametric Statistics,* John Wiley, New York.

Cornell, J. A. (1990). *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data,* 2nd ed., John Wiley, New York.

Cornell, J. A. (1995). "Fitting models to data from mixture experiments containing ordered factors." *Journal of Quality Technology,* vol. 27, pp. 13–33.

Cornell, J. A. (1990*a*). *How to Run Mixture Experiments for Product Quality* (rev. ed.). American Society for Quality Control, Milwaukee.

Cornell, J. A. (1990*b*). *How to Apply Response Surface Methodology* (rev. ed.). American Society for Quality Control. Milwaukee.

Cox, D. R. (1958). *Planning of Experiments.* John Wiley, New York.

Crocker, D. C. (1990). *How to Use Regression Analysis in Quality Control* (rev. ed.). American Society for Quality Control, Milwaukee.

Crosier, R. B. (1984). "Mixture experiments: geometry and pseudo-components." *Technometrics,* vol. 26, pp. 209–216.

Crowder, M. J., and Hand, D. J. (1990). *Analysis of Repeated Measures,* Chapman and Hall, London.

Daniel, C. (1959). "Use of half-normal plots for interpreting factorial two-level experiments." *Technometrics,* vol. 1, pp. 311–341.

Deming, W. E. (1982). *Quality, Productivity and Competitive Position.* MIT Center for Advanced Engineering Study, Cambridge, MA.

Diamond, W. (1989). *Practical Experimental Designs.* Van Nostrand Reinhold, New York.

Dixon, W. J., and Massey, F. J. (1969). *Introduction to Statistical Analysis.* McGraw-Hill, New York.

Draper, N. R. (1985). "Small composite designs." *Technometrics,* vol. 27, pp. 173–180.

Draper, N. R., and Hunter, W. G. (1967). "Transformations: Some examples revisited." *Technometrics,* vol. 11, p. 53.

Draper, N. R., and John, J. A. (1988). "Response surface designs for quantitative and qualitative variables." *Technometrics,* vol. 30, pp. 423–428.

Draper, N. R., and Lin, D. K. J. (1990). "Small response surface designs." *Technometrics,* vol. 32, pp. 187–194.

Draper, N. R., Lewis, S. M., John, P. W. M., Prescott, P., Dean, A. M., and Tuck, M. G. (1993). "Mixture designs for four components in orthogonal blocks." *Technometrics,* vol. 35, pp. 268–276.

Draper, N. R., and Smith, H., Jr. (1981). *Applied Regression Analysis,* 2nd ed., John Wiley, New York.

Duncan, A. J. (1974). *Quality Control and Industrial Statistics,* 4th ed. Richard D. Irwin, Homewood, IL.

Duncan, D. B. (1955). "Multiple range and multiple F tests." *Biometrics,* vol. 11, p. 1.

Dunn, O., and Clark, V. (1987). *Applied Statistics: Analysis of Variance and Regression,* 2nd ed., John Wiley, New York.

Dunnett, C. W. (1964). "New tables for multiple comparison with a control." *Biometrics,* vol. 11, p. 1.

Dyer, D. D., and Keating, J. P. (1980). "On the determination of critical values of Bartlett's test." *Journal of the American Statistical Association,* vol. 75, pp. 313–319.

Easterling, R. G. (1985). "Discussion of Kacker's paper." *Journal of Quality Technology,* vol. 17, pp. 191–192.

Ehrenfeld, S. (1972). "On group sequential sampling." *Technometrics,* vol. 14, pp. 167–174.

Everitt, B. S. (1994). *A Handbook of Statistical Analyses Using S-Plus,* Chapman and Hall, London.

Fisher, R. A., and Yates, F. (1964). *Statistical Tables for Biological, Agricultural, and Medical Research,* 6th ed. Stechert-Hafner, New York.

Fleiss, J. L. (1986). *The Design and Analysis of Clinical Experiments,* John Wiley, New York.

Fries, A., and Hunter, W. G. (1980). "Minimum aberration 2^{k-p} designs." *Technometrics,* vol. 22, pp. 601–608.

Gibbons, J., and Chakraborti, S. (1992). *Nonparametric Statistical Inference,* 3rd ed. Marcel Dekker, New York.

Goh, T. N. (1993). "Taguchi methods: some technical, cultural and pedagogical perspectives." *Quality & Reliability International,* vol. 9, pp. 185–202.

Gorman, J. W., and Hinman, J. E. (1962). "Simplex lattice designs for multicomponent systems." *Technometrics,* vol. 4, pp. 463–487.

Grove, D. M., and Davis, T. P. (1991*a*). *Engineering Quality and Experimental Design,* John Wiley, New York.

Grove, D. M., and Davis, T. P. (1991*b*). "Taguchi's idle column method." *Technometrics,* vol. 33, pp. 349–354.

Gunst, R. F., and Mason, R. L. (1991). *How to Construct Fractional Factorial Experiments,* ASQC Quality Press, Milwaukee.

Haaland, P. D. (1989). *Experimental Design in Biotechnology.* Marcel Dekker, New York.

Haaland, P. D., and O'Connell, M. R. (1995). "Inference for effect-saturated fractional factorials." *Technometrics,* vol. 37, pp. 82–93.

Hahn, G. J. (1977). "Some things engineers should know about experimental design." *Journal of Quality Technology,* vol. 9, pp. 13–20.

Hahn, G. J. (1984). "Experimental design in a complex world." *Technometrics,* vol. 26, pp. 19–31.

Hahn, G. J., and Shapiro, S. S. (1966). "A catalog and computer program for the design and analysis of orthogonal symmetric and asymmetric fractional factorial experiments," Technical Report 66-C-165, G.E. Research & Development Center, Schenectady, NY.

Hayre, L. S. (1985). "Group sequential sampling with variable group sizes." *Journal of the Royal Statistical Society, Series B,* vol. 47, pp. 463–487.

Hicks, C. R. (1982). *Fundamental Concepts in the Design of Experiments,* 3rd ed., Holt, Reinhart & Winston, New York.

Hill, W. J., and Hunter, W. G. (1966). "A review of response surface methodology: a literature survey." *Technometrics,* vol. 8, pp. 571–590.

Hoadley, A., and Kettenring, J. (1990). "Communications between statisticians and engineers/physical scientists." *Technometrics,* vol. 32, pp. 243–274.

Hogg, R., and Ledolter, J. (1992). *Applied Statistics for Engineers and Physical Scientists,* 2nd ed. Macmillan, New York.

Hollander, M., and Wolfe, D. (1973). *Nonparametric Statistical Methods.* John Wiley, New York.

Hunter, J. S. (1959). "Determination of optimum operating conditions by experimental methods," Parts I, II, and III. *Industrial Quality Control,* December 1958, January, February 1959.

Hunter, J. S. (1980). "The national system of scientific measurement." *Science,* vol. 210, pp. 869–874.

Hunter, J. S. (1985*a*). "Statistical design applied to product design." *Journal of Quality Technology,* vol. 17, pp. 210–221.

Hunter, J. S. (1989). "Let's all beware the Latin Square." *Quality Engineering,* vol. 4, pp. 453–466.

Hunter, W. G. (1977). "Some ideas about teaching design of experiments with 2^5 examples of experiments conducted by students." *The American Statistician,* vol. 31, pp. 12–17.

Hurley, P. (1994). "Interactions: Ignore them at your own risk." *Journal of Quality Technology,* vol. 21, pp. 174–178.

Hwang, F. K. (1984). "Robust group testing." *Journal of Quality Technology,* vol. 16, pp. 189–195.

Iglewicz, B., and Hoaglin, D. C. (1993). *How to Detect and Handle Outliers.* ASQC Quality Press, Milwaukee.

Iman, Ronald L. (1994). *A Data Based Approach to Statistics.* Duxbury Press, Belmont, CA.

Ishikawa, K. (1985). *What is Total Quality Control?* Prentice Hall, Englewood Cliffs, NJ.

Jones, M. C., and Rice, J. A. (1992). "Displaying the important features of a large collection of similar curves." *The American Statistician,* vol. 46, pp. 140–145.

Kacker, R. N. (1985). "Off-line quality control, parameter design and the Taguchi method" (with discussion). *Journal of Quality Technology,* vol. 17, pp. 176–209.

Kacker, R. N., and Shoemaker, A. C. (1986). "Robust design: a cost effective method for improving manufacturing process." *AT&T Technical Journal,* vol. 65, pp. 39–50.

Kacker, R. N., and Tsui, K. L. (1990). "Interaction graphs: graphical aids for planning experiments." *Journal of Quality Technology,* vol. 22, pp. 1–14.

Kastembaum, M. A., Hoel, D. G., and Bowman, K. O. (1970). "Sample size requirements; one-way analysis of variances, randomized block designs." *Biometrika,* vol. 57, pp. 421–430, 573–578.

Khuri, A. I., and Cornell, J. A. (1987). *Response Surfaces: Designs and Analyses,* Marcel Dekker, New York.

Ku, H. H. (ed.) (1969). *Precision Measurement and Calibration—Statistical Concepts and Procedures,* National Bureau of Standards Special Publication 300, Vol. 1. U.S. Government Printing Office, Washington, DC.

Kurtz, T. E., Link, B. F., Tukey, J. W., and Wallace, D. L. (1965). "Short-cut multiple comparisons for balanced single and double classifications." *Technometrics,* vol. 7, pp. 95–165.

Lenth, R. V. (1989). "Quick and easy analysis of unreplicated factorials." *Technometrics,* vol. 31, pp. 469–473.

Leon, R. V., Shoemaker, A. C., and Kacker, R. N. (1987). "Performance measure independent of adjustment: an explanation and extension of Taguchi's signal to noise ratio." *Technometrics,* vol. 29, pp. 253–285.

Lilliefors, H. W. (1967). "On the Kolmogorov-Smirnov test for normality with mean and variance unknown." *Journal of the American Statistical Association,* vol. 62, pp. 399–402.

Lin, D. K. J. (1993). "A new class of supersaturated designs." *Technometrics,* vol. 35, pp. 28–31.

Lin, D. K. J. (1995). "Generating systematic supersaturated designs." *Technometrics,* vol. 37, pp. 213–225.

Lucas, J. M. (1985). "Comment on `Off line quality control, parameter design and the Taguchi method' by K. N. Kacker." *Journal of Quality Technology,* vol. 17, pp. 195–197.

Mandel, J. (1964). *Statistical Analysis of Experimental Data.* John Wiley, New York.

Margolin, B. H. (1969). "Results on factorial designs of resolution IV for the 2^n and 2^n3^m series." *Technometrics,* vol. 11, pp. 431–444.

Mason, R. L., Gunst, R. F., and Hess, J. L. (1989). *Statistical Design and Analysis of Experiments with Applications to Engineering and Science.* John Wiley, New York.

Miller, R. G., Jr. (1981). *Simultaneous Statistical Inference,* 2nd ed. Springer-Verlag, New York.

Miller, A., Sitter, R. R., Wu, C. F. J., and Long, D. (1993). "Are large Taguchi-style experiments necessary? A reanalysis of gear and pinion data." *Quality Engineering,* vol. 6, pp. 21–37.

Moen, R. D., Nolan, T. W., and Provost, L. P. (1991). *Improving Quality Through Planned Experimentation.* McGraw-Hill, New York.

Montgomery, D. C. (1991). *Introduction to Statistical Quality Control,* 2nd ed. John Wiley, New York.

Montgomery, D. C. (1991). *Design and Analysis of Experiments.* John Wiley, New York.

Montgomery, D. C. (1991). "Using fractional factorial designs for robust design process development." *Quality Engineering,* vol. 3, pp. 193–205.

Montgomery, D. C., and Voth, S. R. (1994). "Multicollinearity and leverage in mixture experiments." *Journal of Quality Technology,* vol. 26, pp. 96–108.

Mundel, A. (1984). "Group testing." *Journal of Quality Technology,* vol. 16, pp. 181–188.

Myers, R. H., Khuri, A. I., and Carter, W. H. (1989). "Response surface methodology: 1966–1988." *Technometrics,* vol. 31, pp. 137–157.

Myers, R. H., and Montgomery, D. C. (1995). *Response Surface Methodology.* John Wiley, New York.

Nachtscheim, C. J. (1987). "Tools for computer aided design of experiments." *Journal of Quality Technology,* vol. 19, pp. 132–160.

Nair, V. J. (1988). "Analyzing dispersion effects from replicated factorial experiments." *Technometrics,* vol. 30, pp. 247–257.

Nair, V. J., ed. (1992). "Taguchi's parameter design: A panel discussion." *Technometrics,* vol. 34, pp. 127–161.

Nair, V. J., and Shoemaker, A. C. (1990). "The role of experimentation in quality engineering: a review of Taguchi's contributions," pp. 247–277 in *Statistical Design and Analysis of Experiments,* Ghosh, G., ed. Marcel Dekker, New York.

Natrella, M. C. (1963). *Experimental Statistics,* National Bureau of Standards Handbook 91, U.S. Government Printing Office, Washington, DC.

Nelson, L. S. (1983). "Exact critical values for use with the analysis of means." *Journal of Quality Technology,* vol. 15, pp. 40–44.

Nelson, P. R. (1993). "Additional tables for the analysis of means and extended tables of critical values." *Technometrics,* vol. 35, pp. 61–71.

Neter, J., Wasserman, W., and Kutner, M. H. (1990). *Applied Linear Statistical Models: Regression, Analysis of Variance and Experimental Designs,* 3rd ed. Richard D. Irwin, Homewood, IL.

Noether, G. E. *Introduction to Statistics: The Nonparametric Way.* Springer-Verlag, New York.

Odeh, R. E., and Fox, M. (1991). *Sample Size Choice: Charts for Experimenters with Linear Models.* Marcel Dekker, New York.

Oehlert, G. W. (1994). "Isolating one cell interactions." *Technometrics,* vol. 36, pp. 403–408.

Ott, E. R. (1967). "Analysis of means—a graphical procedure." *Industrial Quality Control,* vol. 24, pp. 101–109.

Ott, E. R., and Schilling, E. G. (1990). *Process Quality Control,* 2nd ed. McGraw-Hill, New York.

Owens, D. B. (1962). *Handbook of Statistical Tables.* Addison-Wesley, Reading, MA.

Patel, M. S. (1962). "Group screening with more than two stages." *Technometrics,* vol. 4, pp. 209–217.

Peace, G. S. *Taguchi Methods: A Hands-on Approach.* Addison-Wesley, Reading, MA.

Phadke, M. S. (1989). *Quality Engineering Using Robust Design.* Prentice Hall, Englewood Cliffs, NJ.

Phadke, M. S., Kacker, R. N., Speeney, D. V., and Greico, M. J. (1983). "Off-line quality control in integrated circuit fabrication using experimental design." *Bell System Technical Journal,* vol. 62, pp. 1273–1309.

Piepel, G. F., and Cornell, J. A. (1994). "Mixture experiment approaches: Examples, discussion and recommendations." *Journal of Quality Technology,* vol. 26, pp. 177–196.

Plackett, R. L., and Burman, J. P. (1946). "The Design of Optimum Multi-Factorial Experiments." *Biometrica* vol. 33, pp. 305–325.

Pocock, S. (1983). *Clinical Trials.* John Wiley, New York.

Ramig, P. F. (1983). "Applications of the analysis of means." *Journal of Quality Technology,* vol. 15, pp. 19–25.

Rechtshaffner, R. (1967). "Saturated fractions of the 2^n and 3^m factorial designs." *Technometrics,* vol. 9, pp. 569–575.

Scheffé, H. (1953). "A method for judging all contrasts in the analysis of variance," *Biometrika,* vol. 40, p. 87.

Scheffé, H. (1958). "Experiments with mixtures." *Journal of the Royal Statistical Society, Series B,* vol. 20, pp. 344–360.

Schilling, E. G. (1973). "A systematic approach to the analysis of means." *Journal of Quality Technology,* vol. 5, pp. 93–108, 147–159.

Schmidt, S. R., and Launsby, R. G. (1991). *Understanding Industrial Designed Experiments,* Air Academy Press, Colorado Springs.

Shoemaker, A., Tsui, K.-L., Wu, J. C. F. (1991). "Economical Experimentation Methods for Robust Design." *Technometrics,* vol. 33, pp. 415–428.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components.* John Wiley, New York.

Sheesley, J. H. (1980). "Attributes comparison of k samples involving variables or data using the analysis of means." *Journal of Quality Technology,* vol. 12, pp. 47–52.

Snedecor, G. W., and Cochran, W. G. (1980). *Statistical Methods.* The Iowa University Press, Ames, IA.

Snee, R. D. (1973). "Design and analysis of mixture experiments." *Journal of Quality Technology,* vol. 3, pp. 159–169.

Snee, R. D. (1974). "Computation and Use of Expected Mean Squares in Analysis of Variance." *Journal of Quality Technology,* vol. 6, pp. 128–137.

Snee, R. D. (1979). "Experimental designs for mixture systems with multicomponent constraints." *Communications in Statistics,* vol. A8, pp. 303–326.

Snee, R. D. (1983). "Graphical analysis of process variation studies." *Journal of Quality Technology,* vol. 15, pp. 76–88.

Snee, R. D. (1985). "Computer-aided design of experiments: Some practical experiences." *Journal of Quality Technology,* vol. 17, pp. 222–236.

Snell, E. J., and Simpson, H. R. (1991). *Applied Statistics: A Handbook of GENSTAT Analyses.* Chapman and Hall, London.

Spector, P. (1994). *An Introduction to S and S-Plus.* Duxbury Press, Belmont, CA.

Spurrier, J. D. (1992). "Optimal designs for comparing the variances of several treatments with that of a standard treatment." *Technometrics,* vol. 34, pp. 332–339.

Stephens, Matthew P. (1994). "Comparison of robustness of Taguchi's method with classical ANOVA under conditions of homogeneous variances." *Quality Engineering,* vol. 7, pp. 147–167.

Taguchi, G. (1978). "Off-line and on-line quality control systems." *Proceedings of International Conference on Quality,* Tokyo.

Taguchi, G. (1986). *Introduction to Quality Engineering.* Asian Productivity Association, Tokyo.

Taguchi, G. (1987). *Systems of Experimental Design,* Vol. 1. Unipub, Kraus International Publications, New York.

Tang, K., and Tang, J. (1994). "Design of screening procedures: A review." *Journal of Quality Technology,* vol. 26, pp. 209–226.

Taylor, G. A. R. (1994). "Analysis of experiments by using half normal plots." *The Statistician,* vol. 43, pp. 529–536.

Thomas, G. E., and Kiwanga, S. S. (1993). "Use of ranking and scoring methods in the analysis of ordered categorical data from factorial experiments." *The Statistician,* vol. 42, pp. 55–67.

Thompson, W. O., and Myers, R. H. (1968). "Response surface designs for mixture problems." *Technometrics,* vol. 10, pp. 739–756.

Tribus, M., and Sconyi, G. (1989). "An alternative view of the Taguchi approach." *Quality Progress,* vol. 22, pp. 46–48.

Tukey, J. W. (1949). "Comparing individual means in the analysis of variance." *Biometrics,* vol. 5, p. 99.

Vining, G. C., and Myers, R. H. (1990). "Combining Taguchi and response surface philosophies: A dual response approach." *Journal of Quality Technology,* vol. 22, pp. 38–45.

Wadsworth, H. (ed.) (1990). *Handbook of Statistical Methods for Engineers and Scientists,* McGraw-Hill, New York.

Wang, J. C., and Wu, C. F. J. (1991). "An approach to the construction of asymmetrical orthogonal arrays." *Journal of the American Statistical Association,* vol. 86, pp. 450–456.

Watson, G. S. (1961). "A study of the group screening method." *Technometrics,* vol. 3, pp. 371–388.

Webb, S. R. (1968*a*). "Non-orthogonal designs of even resolution." *Technometrics,* vol. 10, pp. 291–299.

Webb, S. R. (1968*b*). "Saturated sequential factorial design." *Technometrics,* vol. 10, pp. 535–550.

Wernimont, G. (1951). "Design and Analysis of Interlaboratory Studies of Test Methods." *Analytical Chemistry,* vol. 23, p. 1572.

Welch, W. J., Yu, T. K., Kang, K. S., and Sacks, J. (1990). "Computer experiments for quality control by parameter design." *Journal of Quality Technology,* vol. 22, pp. 15–22.

Winer, B. J. (1971). *Statistical Principles in Experimental Design,* 2nd ed. McGraw-Hill, New York.

Wu, C. F. J., and Chen, Y. Y. (1992). "A graph aided method for planning two-level experiments when certain interactions are important." *Technometrics,* vol. 34, pp. 162–175.

Wu, C. F. J., Mao, S. S., and Ma, F. S. (1990). "Sequential elimination of levels—An investigation method using orthogonal arrays" (Chinese). *Chinese Journal of Applied Probability and Statistics,* vol. 6, pp. 185–204.

Youden, W. J. (1959). "Graphical diagnosis of interlaboratory tests." *Industrial Quality Control,* vol. 15, pp. 1–5.

Youden, W. J. (1963). "Ranking laboratories by round-robin tests." *Materials Research and Standards,* vol. 3, pp. 9–13.

Youden, W. J. (1967). *Statistical Techniques for Collaborative Tests.* Association of Official Analytical Chemists, Washington, DC.

Youden, W. J., and Hunter, J. S. (1955). "Partially replicated Latin squares." *Biometrics,* vol. 5, pp. 99–104.

Zahn, D. A. (1975). "Modifications of and revised critical values for the half-normal plot." *Technometrics,* vol. 17, pp. 189–200.