CHAPTER 85
Design of Experiments

H. SAMUEL WANG
Chung Yuan Christian University

CHUNG-PU CHANG
Eureka Consulting Co.

1. INTRODUCTION 2225
1.1. Perspective 2225
1.2. Statistical Experiments 2225
1.3. Basic Definitions 2225

2. PLANNING FOR EXPERIMENTS 2226
2.1. Program and Activities: Steps and Checkpoints 2226
2.1.1. Stage 1: PLAN 2226
2.1.2. Stage 2: DO 2226
2.1.3. Stage 3: STUDY 2227
2.1.4. Stage 4: ACT 2227
2.2 Size of Experiments 2227

3. GOOD EXPERIMENTAL PRACTICES 2228
3.1. Randomization 2228
3.2. Blocking 2228
3.3. Replication 2228

4. PRECAUTIONS FOR EXPERIMENTAL DESIGNS 2228
4.1. Ten Commandments for Experimental Designs 2228

5. FUNDAMENTAL DESIGNS AND CONCEPTS 2229
5.1. A Case: Weight Watch Program 2229
5.2. Fixed-Effect and Random-Effect Models 2229
5.3. Completely Randomized Design (CRD) 2230
5.4. Randomized Complete Block Design (RCBD) 2230
5.5. Latin Square Designs and Interactions 2230
5.6. Factorial Design 2230
5.7. 2^k and 3^k Factorial Designs 2231
5.8. Fractional Factorial Designs 2231
5.9. Orthogonal Arrays 2232

6. ANALYSIS OF A BASIC DESIGN 2232
6.1. Hypotheses and Models 2232
6.2. ANOVA: Analysis of Variance 2233
6.3. Marginal Averages 2234
6.4. Rationale of ANOVA Analysis 2234

7. SCREENING DESIGNS 2235
7.1. Strategy of Screening Design 2235
7.2. Weight Watch Experiment Using $L_8(7)$ 2235
7.3. ANOVA 2235
7.4. Recommendations 2236

8. PARAMETER DESIGNS 2237
8.1. Strategy of Parameter Design 2237
8.2. Concepts of Parameter Design 2237
8.3. Weight Watch Experiments 2238

9. THE STRATEGIES OF EXPERIMENTS 2238

10. CONCLUSION 2239

REFERENCES 2239

ADDITIONAL READING 2240

1. INTRODUCTION

1.1. Perspective

Experimentation is common in every aspect of life. As part of a problem-solving program, experiments are carried out in order to observe the effects of changes under a controlled framework. Through one or more iterations of experiments, adequate knowledge is acquired or confirmed. Know-how, sometimes coupled with know-why, is gathered and used for decision making. Experiments are indispensable to the learning process.

The need for learning through experiments is particularly obvious in industry, whether in manufacturing or the service sector. Consider the development and marketing of a new drug. After a new drug is found to be effective for treating a certain kind of cancer, a series of experiments is usually conducted before the drug is formally marketed. For instance, a laboratory scientist performs experiments to identify other supportive constituents. With the aid of these experimental results, he or she picks the most effective composition. A manufacturing engineer uses experiments to determine process conditions such as pressure, temperature, flow rate, the catalyst quantity, and so forth. Thus the goal of fabricating quality medicine at the lowest possible cost is achieved. A marketing staff relies on computer simulation, which in fact is a form of numerical experimentation, to pick the most profitable marketing strategy. An FDA officer carries out experiments to detect potential adverse effects on different consumers determined by age, sex, and ethnicity. Experimentation is also important to the consumer organization. It is relied upon to compare the effectiveness of this new drug against others existing in the market.

Other examples illustrating when and where experiments are performed are numerous (Diamond 1989, 1997; John 1998).

1.2. Statistical Experiments (Box et al. 1978; Du Pont Co. 1988)

An experiment is often confused with a trial or a test, which in practice takes no account of experimental errors due to inherent variations. In fact, variations occur in every component and stage of experimentation, including variation in experiment parameters, due to inaccurate setting of machines and instruments, in methods and handling, in measurements, and due to analysis.

Moreover, experiments are often run by intuition with factors varied one at a time. This is not only ineffective costwise, it also causes the risk of reaching incorrect conclusions due to negligence of the potential interactions among factors

In the following, we are concerned with the design and analysis of experiments based on statistical considerations. These are often referred to as statistical experiments.

A statistical experiment serves as a means to compare and choose the most effective treatment, identify significant factors, reveal cause-and-effect mechanisms, and determine optimal process conditions. It therefore plays an important role in quality improvement, productivity increases, cost saving, and management decisions.

In essence, a statistical experiment implies a systematic varying of process, observation of change in response, collection and analysis of data, and extraction of information to arrive at a conclusion. Experiments are designed so that the appropriate decision can be arrived at in the shortest time and within cost constraints.

1.3. Basic Definitions (Anderson and McLean 1974; Du Pont Co. 1988; Hunter 1998; Montgomery 1996)

An experimental design is a formal plan for execution of the experiment. It includes the choice of response, factors, designation of levels, and assignment of blocks as well as application of treatment on experimental units.

These commonly used terms are defined and explained below:

- *Response:* A response is the dependent variable that corresponds to the outcome or resulting effects of interest in the experiment. One or more response variables may be studied simultaneously.
- *Factors:* A factor is a variable contribute to the response. A factor can be controllable or uncontrollable. It may be quantitative, such as pressure in psi or duration time in minutes. It may be qualitative, such as different methods, different operators, or different suppliers.
- *Levels:* The levels are the chosen conditions of the factor under study. They may be quantitative values such as 5%, 8%, and 10% alcohol concentration. They also take categorical forms such as supplier A, B, C, and D.
- *Blocks:* A block is a homogenous portion of the experimental environment or materials that bears certain variation effects on the response(s). A block may be a batch of material supplied

by a vendor or products manufactured in a shift on a production floor. The term *block* is sometimes associated with *factor* and called *block variable*.

- *Treatments:* A treatment is the condition or a factor associated with a specific level in a specific experiment.
- *Experimental units:* Experimental units are the objects or entities that are used for application of treatments and measurements of resulting effects.

2. PLANNING FOR EXPERIMENTS (Du Pont Co. 1988; Hunter 1998)

A complete experimental process involves five stages: including (1) design, (2) data collection, (3) data analysis, (4) interpretation of results, and (5) communication of results. In order to achieve an effective and efficient experimentation, it is of utmost importance to take effort to work out a comprehensive experimental plan.

In the planning stage of the experiment, the design of the experiment implies the careful and thorough consideration of the following issues:

- Global environment of the problem
- Objectives of the study
- Properties to be studied
- Variables to be controlled
- The environment of concern
- Size of experimental units
- Number of experimental runs
- Conduct of the experiments
- Approach for data analysis

Each of the technical, statistical, and administrative aspects of experiments are to be taken into consideration.

2.1. Program and Activities: Steps and Checkpoints (Du Pont Co. 1988; Hunter 1998)

2.1.1. Stage 1: PLAN

1. *Problem Recognition*
  (a) Formation of task force
  (b) Evaluation of strengths, weaknesses, opportunities, and threats
  (c) Identification of problem area(s)

2. *Statement of problem and objective:*
  (a) Determination of ultimate goal(s)
  (b) Determination of en route objectives
  (c) Determination of immediate objectives
  (d) Identification of the cost and time constraints

3. *Design of experiment:*
  (a) Definition of experimental units
  (b) Determination of response variable(s)
  (c) Selection of factors
  (d) Determination of factor levels
  (e) Appraisal of possible interaction
  (f) Choice of design
  (g) Definitions of data and effect models
  (h) Decision of number of replicates
  (i) Setup of execution plan (timetable, sequence schedule, facilities allocation)
  (j) Setup of data-collection plan

2.1.2. Stage 2: DO

1. *Mission orientation:*
  (a) Explanation of objectives and tasks to be achieved
  (b) Attention to precautions in execution and data recording

2. *Physical preparations:*
   (a) Preparation of experimental units
   (b) Development of methods and facilities needed
3. *Execution of experiments:*
   (a) Execution in accordance with prescribed conditions
   (b) Control of schedules
   (c) Varying and control of treatments in terms of randomization
4. *Observation and measurement:*
   (a) Surveillance of condition of facilities
   (b) Control of measurement procedures
5. *Recording:*
   (a) Recording of program by date, run numbers, etc.
   (b) Recording of abnormal situations
   (c) Recording of change of designs
   (d) Recording of measured data
   (e) Recording of data from extra or missing experiments

2.1.3 Stage 3: STUDY

1. *Analysis of data:*
   (a) Diagnosis of data
   (b) Application of appropriate statistical methods
   (c) Graphical analysis
   (d) Checking of model adequacy
2. *Interpretation of results:*
   (a) Identification of substantial factors
   (b) Selection of desired levels
   (c) Estimation of factor effects
   (d) Account for limitations in data acquisition or analysis
   (e) Interpretation in terms of statistical, technical, and economical significance

2.1.4. Stage 4: ACT

1. *Confirmation of conclusion*
2. *Presentation of results:*
   (a) Preparation of report
   (b) Use of graphical and tabular forms
   (c) Indication of implications for potential applications
3. *Recommendations:*
   (a) For process change
   (b) For further experiments
   (c) For modification in strategies of experiment
4. *Process change and standardization*
5. *New situation appraisal*
6. *Preparation for further experiments*

2.2. Size of Experiments

Each experimenter should be concerned with the size of the experiment. A large enough experiment enables the detection of the significant effects of the factors in the response and thus ensures obtaining some know-how from the study. Yet it must be small enough to ensure that the cost of the experimentation is within the allocated budget and can be completed within the assigned time frame.

The number of experimental runs or experimental units is to be determined beforehand. Its precise determination involves statistical computation that requires prestated probability of committing type I and II errors, the desired accuracy in detecting the difference between the means of the responses resulting from different treatments. Besides, other statistical parameters are also required. As a rule of thumb, the appropriate size for an experiment is between 8 and 60. For more elaborate evaluation, consult statistical handbooks (Box et al. 1978; Daniel 1976; Wadsworth 1990; Winer et al. 1991).

3. GOOD EXPERIMENTAL PRACTICES (Hicks 1982; Montgomery 1996)

In the process of experimentation, there exist two types of errors: random errors and bias errors. Random error is experimental error for which the numerical values change from one run to another without a consistent pattern. It can be thought of as inherent noise in measured responses. Bias error is experimental error for which the numerical values tend to follow a consistent pattern over a number of experimental runs. It is attributed to an assignable cause. To reduce the effects of both types of errors, it is strongly adviced that the following good experimental practices be taken into consideration.

Replication, randomization, and blocking, the three measures for ensuring a successful experiment as presented below, are called the three Fisherean principles of experimental designs. They are attributed to R. A. Fisher, the forerunner of the modern design of experiments.

3.1. Randomization

Randomization is the procedure of assigning the experimental units to various treatments in a purely chance manner. It is also used for arrangement of the experiments in random order. It is intended to balance out the effect of uncontrollable variables. Because bias errors are not confused with the effect of the factors, the quality of data is improved and statistical inferences are made possible.

To achieve randomization, the order of experiment is scrambled so that any bias present will be mixed up and become a part of the random variation. One of the following options can be taken

The trial numbers can be written on small slips of paper and selected at random.

A table of random numbers can be employed to assign a run order and the trials.

3.2. Blocking

The source of bias error in an experiment may accompany differences among blocks, namely batches of raw materials, production machine, hours within a day, or seasons of the year. It is necessary to reduce their influence by proper design of the experiment. Blocking means running the experiment in a specially chosen subgroup that allows removal of the effect of bias errors that are confounded with the main factors.

To achieve blocking, the run order is broken up into smaller units so that the bias is negligible within the block. Note that under these circumstances a separate randomization is needed in each block.

3.3. Replication

Replication involves the repetition of experimental runs so that more than one observation for each treatment combination is available for statistical analysis.

The benefit of replication is that the average of several observations comes closer to the true value than a single observation. Replication helps balance out the bias due to the effect of nuisance factors. It also helps to detect gross errors in the measurements. It therefore improves the precision of the statistical inferences.

Different randomization applies to different replications of the experiment.

4. PRECAUTIONS FOR EXPERIMENTAL DESIGNS

4.1. Ten Commandments for Experimental Designs (Gryna and Juran 1993; Montgomery 1996)

1. *Don't set out without a clear problem definition and objective statement:* An unplanned experiment often ends up in total loss of time and money.
2. *Do keep the design and analysis of experiment as simple as possible:* A comprehensive but complicated design of experiment may cost much more and end nowhere.
3. *Don't rely solely on statistical experts:* Interaction with subject matter specialists makes professional insight invaluable.
4. *Don't underestimate the importance of randomization:* The adverse effect of systematic errors can be of vital importance.
5. *Don't start statistical analysis without first challenging the validity of the data:* What can you expect out of garbage input?
6. *Don't throw away outliers without solid reasoning:* Every piece of data stores a hidden story waiting to be opened.
7. *Do make full use of graphical presentations:* A picture can be worth more than hundreds of words.

TABLE 1 A Balanced Design

DIET	D_1	D_2	D_3
	y_{11}	y_{21}	y_{31}
	y_{12}	y_{22}	y_{32}
	y_{13}	y_{23}	y_{33}

8. *Do avoid statistical jargon in conclusion and report writing:* Problem language is the only thing that is universal in a corporation.
9. *Don't blindly follow statistical conclusions without taking into account their practical significance and economic considerations:* Negligence of the nonstatistical aspects of the experimental design can prove to be vital.
10. *Don't think that one iteration of a time experiment can solve the problem once and for all:* The outcome of one experiment often provides a direction for further iterations of exploration.

5. FUNDAMENTAL DESIGNS AND CONCEPTS (Anderson and McLean 1974; Box and Draper 1987; Gryna and Juran 1993; Hunter 1998; Montgomery 1996)

Statistical experiments can have various objectives and constraints. Identifying the most influential factor(s) or independent variables and their respective effect on the response or dependent variable(s) is one of the most common objectives. The nature and number of factors of interest, the number of levels a factor can vary, the limitation in time span and budget, and so forth are some of the common constraints. Depending on its specific objective and constraints, an experiment can have various designs. In the following we take a weight-watch program as an example to illustrate various alternative designs and their underlying principles.

5.1. A Case: Weight-Watch Program

One response variable of concern is the weight loss (WTLOSS) in kilograms measured in the first month. The main factor DIET has three levels, that is, options, D_1, D_2, and D_3. Suppose nine persons volunteer to participate in this experiment program. One natural arrangement is to assign three persons to each of the three DIETs, as shown in Table 1. Note that this is a balanced design in the sense that the same number of persons received the same treatment. For a simple design like this, balancing is not a must but is desired. However, in other cases it is almost a requirement for easy analysis and guarantees the same degree of precision in estimation of different treatments.

Under this arrangement, one diet is assigned to three volunteers, that is, each has three replications. Replication is used to estimate the experimental error. It also helps to increase the precision.

5.2. Fixed-Effect and Random-Effect Models

In this experiment, the three DIETs may be the only ones of concern to the experimenter. In this case it is called a fixed-effect model, and the conclusion drawn is applicable only to these specific three options. However, situations may arise in which the experimenter is seeking for a conclusion applicable to all possible DIET options and yet he or she can only handle three options. Then, three of the available options should be drawn from all possible at random. This is called a random-effect model. In addition to the difference in the scope of conclusion, different types of effect models also imply a somewhat different approach of data analysis. In what is to follow, only the fixed-effect model is considered.

An advertent error can easily occur at this point. Very often, experimental units are assigned to treatments by convenience. For example, the WTLOSS may vary from one ethnic group to another. Yet persons of the same ethnic group (ETHNIC) are assigned the same diet. One extreme case could appear like the one shown in Table 2, where all Caucasians (C) are assigned to D_1, all Africans (A)

TABLE 2 Confounding of Factors

DIET	D_1	D_2	D_3
	C	A	O
	C	A	O
	C	A	O

to D_2, and all Orientals (O) to D_3. One problem arises: Whenever one of the DIETs is found to be more effective than the others, we cannot separate the confounding of the two factors and thus can hardly claim which is the truly dominant one.

5.3. Completely Randomized Designs (CRD)

To overcome this problem, the experimental units are usually assigned to treatments by randomization, that is, in a purely chance manner. This way of completely randomized allocation of DIETs to volunteers, as shown in Table 3, is called completely randomized design (CRD).

5.4. Randomized Complete Block Design (RCBD)

Another way to guard against any possible bias due to the effect of ETHNIC is to carry out the experiment by the randomized complete block design (RCBD), as shown in Table 4. Here, each of the three ethnic groups constitutes a block and receives all of the three treatments in random order.

A RCBD has the advantage of eliminating the contamination of the block factor on the main factor. It permits the removal of the block effect from the experimental error and thus provides a more decisive conclusion. Moreover, the effect of the block factor can usually be tested and evaluated.

5.5. Latin Square Designs and Interactions

At this point one may suspect that different amount of EXERCISE, say light (L), medium (M), and heavy (H), may have different degrees of effect on WTLOSS and should be treated as another block factor. Under these circumstances, the experiment is often carried out according to the Latin square design, as shown in Table 5. Note that each DIET is assigned only once to each ETHNIC and only once to each EXERCISE. It enables the evaluation of three factors with only nine observations. However, it requires that no interaction exist between the factors.

Two factors are said to have interaction when the effect of one factor varies under the different levels of another factor. The concept of interaction is illustrated in Figure 1.

5.6. Factorial Designs

Whenever interaction exists between factors, the experiment must be run according to the factorial design (FD) shown in Table 6 to ensure accurate and precise conclusion. Note that this experiment covers each of the 27 combinations of the levels of the three factors.

One may note that the factorial design differs from the conventional one-factor-at-a-time approach. In the latter the experiment is run in several iterations, and in each iteration only one factor is varied while the others are held constant. As a result, the factorial design allows the test and evaluation of

TABLE 3 Completely Randomized Design (CRD)

DIET	D_1	D_2	D_3
	W	A	A
	B	B	A
	W	B	W

TABLE 4 Randomized Complete Block Design (RCBD)

DIET	D_1	D_2	D_3
	C	A	A
	A	O	C
	O	C	O

TABLE 5 Latin Square Design

DIET	B	W	A
	D_1	D_2	D_3
	D_2	D_3	D_1
	D_3	D_1	D_2

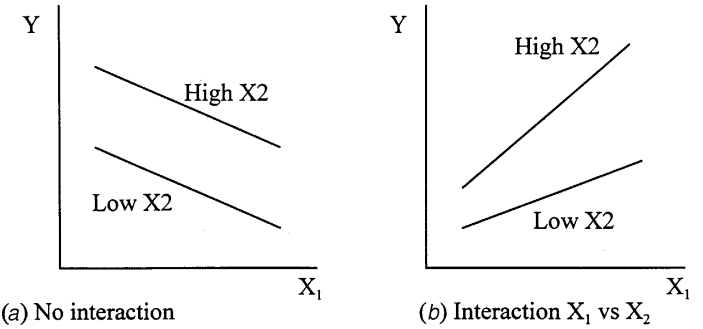

Figure 1 Concept of Interaction.

the effects of all three factors at one iteration. It is therefore more effective in the sense that it requires a shorter time to reach a complete conclusion. It also guards against the risk of missing the optimum of a surface, as is often seen in other approaches.

The FD also has disadvantages. The number of observations increases exponentially with the number of factors and also with their number of levels.

5.7. 2^k and 3^k Factorial Designs

As long as the experiment is still in its exploratory stage and one is mainly interested in screening out the less effective factors, the preceding problem can be partially solved by reducing the number of levels of each factor to two or three. Under such arrangement, the k-factor FD is then denoted as 2^k or 3^k FD, respectively. Although the two-level FD (see Table 7) requires a lower number of observations than its three-level counterpart, it is applicable only when it is believed that none of the factors has a nonlinear effect on the response.

5.8. Fractional Factorial Designs (FFD)

Further reduction of the number of observations can be achieved by employing the fractional factorial design (FFD). The notation 2^{k-p} is used to denote a 2^p fraction of a 2^k fractional design. See Table 8 for the one-half fraction of the 2^3 FD of the weight-watch experiment.

TABLE 6 Factorial Design

DIET (1)		D_1			D_2			D_3		
EXERCISE (2)		L	M	H	L	M	H	L	M	H
ETHNIC (3)	C	y_{111}	y_{121}	y_{131}	y_{211}	y_{221}	y_{231}	y_{311}	y_{321}	y_{331}
	O	y_{112}	y_{122}	y_{132}	y_{212}	y_{222}	y_{232}	y_{312}	y_{322}	y_{332}
	A	y_{113}	y_{123}	y_{133}	y_{213}	y_{223}	y_{233}	y_{313}	y_{323}	y_{333}

TABLE 7 Factorial Design (FD)

Run	DIET	ETHNIC	EXERCISE
1	1	1	1
2	2	1	1
3	1	2	1
4	2	2	1
5	1	1	2
6	2	1	2
7	1	2	2
8	2	2	2

TABLE 8 Fractional Factorial Design (FFD)

Run	DIET	ETHNIC	EXERCISE
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1

TABLE 9 The Orthogonal Array $L_8(2^7)$

Factor Interaction	a	b	$a \times b$	c	$a \times b$	$b \times c$	$a \times b \times c$
Column	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

5.9. Orthogonal Arrays (OA) (Taguchi 1986)

For each FD there exist several alternative FFDs. One specific class of FFDs is called an orthogonal array (OA), often referred to as the Taguchi method. Reduced to its simplest level in allocating factor-level combinations, it has become increasingly popular.

Depending on the number of levels and the number of rows, a cluster of different OAs is available from which to choose.

Shown in Table 9 is one typical two-level OA with eight rows and seven columns, denoted as $L_8(2^7)$. The subscript denotes the number of rows or the number of factor-level combinations this OA provides. The lower number between the parentheses stands for the number of factors it should have, while its superscript stands for the number of columns this OA has or the maximum number of factors the experimenter is allowed to allocate.

Inside the array, numbers 1 and 2 label the low and high levels of the factors. However, each of the rows indicates the factor-level combination to be applied to the specific experimental unit. Note that in each column, both levels 1 and 2 appear an equal number of times. It is also true for all combinations (1,1), (1,2), (2,1) (2,2) appearing in any two columns. Hence the name ''orthogonal array.''

The OA is a balanced design in nature. As a form of fractional factorial design, it has an economic advantage, and it allows evaluation of main effects and the interactions between factors as well.

Attached to the bottom of the table is the component row, which tells how the main effects and interaction effects are associated with columns. In other words, if factors A, B, and C are assigned to columns, 1, 2, and 4, the component row points out that the interactions A X B, A X C, B X C, and A X B X C, are associated with columns 3, 5, 6, and 7. It can therefore be used as an aid to allocate factors to the array.

Besides $L_8(2^7)$, other OAs of the two-level family include $L_4(2^3)$, $L_{16}(2^{15})$, and $L_{32}(2^{31})$. Their counterparts in the three-level family include $L_9(3^4)$ and $L_{27}(1^3)$.

A set of linear graphs is available for each of the OAs to facilitate allocation of factors and interactions to columns of the arrays.

6. ANALYSIS OF A BASIC DESIGN (Box et al. 1978; Hicks 1982; Montgomery 1996)

6.1. Hypotheses and Models

In the weight-watch experiment we are concerned with the problem of comparing the effects of three diets. The hypotheses under test are therefore

H_0: $\mu_1 = \mu_2 = \mu_3$.
H_1: At least one of the three is different from the others.

The true mean of any one treatment can, in fact, be looked upon as the sum of a grand mean μ and the specific effect of the ith treatment α_i, that is, $\mu_i = \mu + \alpha_i$.
The hypotheses under test hence become (Box et al. 1978):

H_0: $\alpha_1 = \alpha_2 = \alpha_3 = 0$.
H_1: At least one of α_i's is unequal to zero.

Realizing the existence of variation due to environment, experimental units, execution process, and also measurement errors, any measured WTLOSS taken from the ith DIET group is unlikely equal to its treatment mean, and is usually decomposed as (Bhote 1991):

$$\begin{aligned} Y_{ij} &= \mu_i + \varepsilon_{ij} \\ &= \mu + \alpha_i + \varepsilon_{ij} \end{aligned}$$

This equation is usually referred to as the data model. The data model of the measured WTLOSS data acquired under the CRD setup shown in Table 3 can be represented as

$$\begin{matrix} \text{Observation} \end{matrix}$$

$$\begin{bmatrix} 6.53 & 3.23 & -0.11 \\ 6.72 & 2.19 & 0.35 \\ 3.91 & 4.72 & 2.61 \end{bmatrix} = \begin{bmatrix} 4 & 4 & 4 \\ 4 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} + \begin{bmatrix} 1.53 & -0.77 & -3.11 \\ 1.72 & -1.81 & -2.65 \\ -1.09 & 0.72 & -0.039 \end{bmatrix}$$

Mean Effect Error

To enable valid statistical analysis, the error component ε_{ij} is required to follow independent identical normal distribution of the mean zero and common variance, or iid $N(0, \sigma^2)$.

6.2. ANOVA: Analysis of Variance

The test of the hypotheses uses the analysis of variance (ANOVA) approach, which is based on the decomposition principle of sum of squares. In other words, the variation of the observations from the grand average can be decomposed into two components: the variation around their group average, or the within-group variation, and the variation between the group average, or the between-group variation.

If the between-group variation is larger than what is expected from the variation that occurs within the groups, we would suspect that group means μ_1, μ_2, μ_3 are not the same. The F distribution is used for checking this point.

The F statistic is computed from

$$F = \frac{\text{SSB}/(k-1)}{\text{SSW}/(N-k)}$$

where SSB = sum of squares due to between-group variation.

$$= \sum_{i=1}^{k} \sum_{j=1}^{n} (\overline{Y}_{ij} - \overline{Y})^2$$

where SSW = sum of squares due to within-group variation.
= SSTO − SSB

$$\text{SSTO} = \sum_{i=1}^{k} \sum_{j=1}^{n} (\overline{Y}_{ij} - \overline{Y})^2$$

where k = number of groups
n = number of observations in each group
$N = kn$

The whole computation procedure is usually summarized in an ANOVA table, as shown in Table 10.

The ANOVA table is provided by almost every statistical software.

TABLE 10 ANOVA Table for WTLOSS Data in CRD

Source	SS	*df*	MS	F_s
Between DIETs	34.13	2	17.07	8.25
Within DIETs	12.41	6	2.07	
Total	46.54	8		

The computed F statistic F_s is now compared to its corresponding critical value at the specified significance level $\alpha = 0.05$, that is, $F_{0.05}(2,6) = 5.14$. Since $F_s > 5.14$, we accept H_1 and conclude that at least one of the DIETs is unequal to the others.

6.3. Marginal Averages

By comparing the marginal averages (see Figure 2) of the three DIETs, those of DIET D_1 have an average 5.72, which is significantly larger than DIETs D_2 and D_3. We are therefore assured that D_1 is the most effective program for a keen weight watcher.

6.4. Rationale of ANOVA Analysis

According to the statistical theory, the mean squares SSB/(k − 1) and SSB/(N − k) have their respective expected mean squares (EMS) as (Box et al. 1978):

$$\text{E(MSB)} = \sigma^2 + n\sigma_A^2$$

$$\text{E(MSE)} = \sigma^2$$

where

$$\sigma_A^2 = \sum_{i=1}^{k} \alpha_i^2/(\text{k} - 1) \text{ for the fixed-effect model}$$

The test of the hypotheses can also be understood by viewing F approximately, as the ratio of $\sigma^2 + \sigma_A^2$ to σ^2. If all α_i's are the same and equal to zero, the ratio E(MSB)/E(MSW) is likely to be close to 1 and therefore leads us to accept H_0. On the other hand, if one of the α_i's is not equal to the others, the same ratio is likely to take a quantity much greater than 1 and therefore leads us to accept H_1.

For the applicable occasions and the analysis of some basic designs, please refer to the summary that follows.

Whenever there are more than one source or variation to be compared to, the EMS can be used as a quick decision aid. As a principle, a designate is always compared with a shorter row having common elements but one.

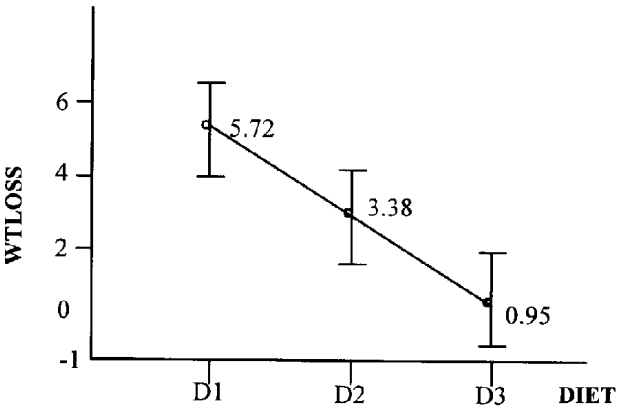

Figure 2 Marginal Average Chart with 95% Confidence Interval.

7. SCREENING DESIGNS (Barker 1990; Box et al. 1978; Dean and Voss 1999)

7.1. Strategy of Screening Design (Du Pont Co. 1988)

Many of the factors initially considered in the early stage of an experimental project may have little or no effect on response. The purpose of a screening experiment is to reduce experimental time and cost by identifying the factors that deserve thorough investigation in the subsequent stages. Therefore, in designing an experiment with many factors, it is useful to start with a screening experiment before going on to the more in-depth studies described in the preceding section.

7.2. Weight-Watch Experiment Using $L_8(2^7)$

Let us return to the weight-watch experiment to illustrate how an OA such as $L_8(2^7)$ is used to assign the factor levels. Suppose the factors and their levels of concern at this stage are as listed here.

Factor	Level 1	Level 2
DIET	DIET 1(D_1)	DIET 2(D_2)
EXERCISE	Medium (M)	High (H)
SEX	Female (F)	Male (M)
ETHNIC	White (W)	Black (B)

Suppose it is anticipated that interaction between DIET and EXERCISE and between DIET and SEX may exist. With the aid of the component row of the $L_8(2^7)$, we choose to assign the factors DIET, EXERCISE, SEX, and ETHNIC to columns 1, 2, 4, and 6. The treatment conditions can then be readily read from the array and are shown in Table 11.

Taking the randomization principle into account, the treatment condition D_1LFW, which is decoded as DIET 1, low exercise on a white female, and so on should be assigned in a random manner.

7.3. ANOVA

Table 11 also serves as a worksheet for data analysis. Recorded in the far-right column of Table 11 are the WTLOSS data resulting from the assigned treatment combinations. The T_1 of any column stands for the total of all observed data that are associated with level 1. Likewise, T_2 stands for the same types of data for level 2. For example, the T_1 for the DIET column is obtained by adding the four observations corresponding to the l's in column 1. Hence,

$$T_1 \text{ (DIET)} = 6.21 + 8.82 + 2.23 + 4.85 = 22.11$$

The task of subsequent analysis is to test the null hypotheses that different levels of the factors DIET, EXERCISE, SEX, and ETHNIC do not have any effect on WTLOSS. It is the same as in analyzing the CRD data; the ANOVA is used to test the previously mentioned hypotheses. The

TABLE 11 Orthogonal Array for Weight-Watch Experiment

Factor Interaction	DIET	EXR	$D \times E$	SEX	$D \times S$	ETH			
Column	1	2	3	4	5	6	7	Treatment Combination	Observed WTLOSS
1	1	1	1	1	1	1	1	D_1MFW	6.21
2	1	1	1	2	2	2	2	D_1MMB	8.82
3	1	2	2	1	1	2	2	D_1FMB	2.23
4	1	2	2	2	2	1	1	D_1HMW	4.85
5	2	1	2	1	2	1	2	D_2MFW	1.06
6	2	1	2	2	1	2	1	D_2MMB	2.75
7	2	2	1	1	2	2	1	D_2HFB	2.75
8	2	2	1	2	1	1	2	D_2HMW	9.26
T_1	22.11	18.84	27.04	12.25	20.45	21.38	16.56		
T_2	15.82	19.09	10.89	25.68	17.48	16.55	21.37		
SS	4.946	0.008	32.602	22.545	1.103	2.916	2.892		

TABLE 12A ANOVA for Weight-Watch Experiment

Source	SS	DOF	MS	F_c
DIET	4.95	1	4.95	1.71
EXERCISE	0.01	1	0.01	0.00
SEX	22.54	1	22.54	7.80
ETHNIC	2.92	1	2.92	1.01
Interaction				
DIET × EXR	32.60	1	32.60	11.27
DIET × SEX	1.1	1	1.10	0.38
ERROR	2.87	1	2.87	
TOTAL	67.01	7		

$F_{0.05}(1,1) = 161$

variation due to the effect of a specific factor is compared to the variation due to random error. The sum-of-squares for the ith column and the corresponding F statistic are computed by

$$\mathrm{SS}_i = \frac{(T_1 - T_2)^2}{8}$$

$$F = \frac{SS_i/df_i}{SSE/df_E}$$

where SSE, the sum-of-squares due to errors, is found by summing up the SS's of the undesignated columns. For the weight-watch experiment, SSE = SS_7. The ANOVA table obtained through this method is shown in Table 12A.

The computed F statistics are now compared to their common critical value at the specified significance level, say, $\alpha = 0.05$, that is, $F_{0.05}(1,1) = 161$. Since none of these is greater than 161,we do not have sufficient evidence to conclude that any of the main effects and interactions are significant at $\alpha = 0.05$. However, the computed F statistics do reveal that while the interaction DIET × EXR is relatively larger than the others, the factor SEX and another interaction DIET × SEX are negligible and thus can be merged with the error term. The resulting ANOVA table is listed in Table 12B.

As the computed F statistics of the interaction DIET × EXR, that is $F_c = 14.15 > 10.128 = F_{0.05}(1,3)$, we do have sufficient evidence to claim that two DIETs do affect WTLOSS differently at the two different levels of EXERCISE amount. It also suggests that SEX may have different WTLOSS. Further examination of the marginal average chart (see Figure 3) reveals that the male under this experimental setup attains greater WTLOSS than the female. It is noted that DIET 1 has effect on WTLOSS when it is used with a medium amount of EXERCISE. However, DIET 2 also produces a sizable WTLOSS when it is used with a higher amount of EXERCISE.

7.4. Recommendations

As mentioned earlier, the OA mainly serves the purpose of screening the vital few from the trivial many. In order to reach a solid and meaningful conclusion, more thorough confirmatory experiments are required. Take the weight-watch experiment as an example. It is recommended that the focus be

TABLE 12B ANOVA for Weight Watch Experiment

Source	SS	DOF	MS	F_c
DIET	4.95	1	4.95	2.15
EXERCISE	0.01	1	0.01	0.00
SEX	22.54	1	22.54	9.79
Interaction				
DIET × EXR	32.60	1	32.60	14.15
ERROR	6.91	3	2.30	
TOTAL	67.01	7		

$F_{0.05}(1,3) = 10.128$

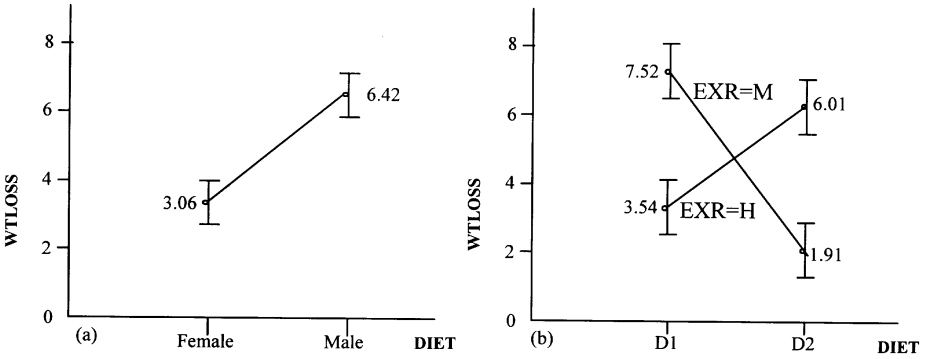

Figure 3 Marginal Average Chart with 95% Confidence Level.

placed at this point on only DIET and EXERCISE as the two main factors. Adding a few more levels may be worthwhile. A replicated two-way factorial design run on both sexes may be the experimenter's best choice. Some statistical tools, such as multiple comparison, confidence interval estimation, residual analysis, normality check, and response surface methodologies, are commonly used.

8. PARAMETER DESIGN (Peace 1989; Phadke 1989; Taguchi and Wu 1980; Taguchi 1986, 1987)

8.1. Strategy of Parameter Design

Traditionally, when performing an experimental design, the experimenter places his or her focus on finding the product or process condition that yields the best mean performance. However, this approach may not satisfy the demands of modern marketing strategies. This strategy requires a robust product or process that satisfies a wide range of customer interests.

For example, the combination of DIET I and the medium exercise amount in the preceding weight-watch experiment may, in fact, yield the greatest amount of WTLOSS, though this conclusion is valid for only some combinations of SEX and ETHNIC. However, a weight-watch service salesperson is more interested in a program that is robust or resistant to noise factors such as SEX and ETHNIC, that is, a program that is suitable for all possible combinations of SEX and ETHNIC.

The same is true for a production engineer. In designing a product or process, in addition to the major parameter settings, noise factors such as manufacturing variation, component tolerance, customer use conditions, and product deterioration need to be taken care of. A good product or process is one that is robust to variations due to these noise factors.

It is this requirement for robustness that prompted Taguchi to develop the concept of parameter design and produced a great impact on the world of experimental design.

8.2. Concepts of Parameter Designs (Barker 1990; Taguchi and Wu 1980; Taguchi 1986)

In the context of parameter design, the simple response is no longer of major interest. Rather, a composite performance measure that integrates both the mean and variance of the response plays the role of a dependent variable. Depending on the nature of the problem under study, various performance measures are developed for situations including "the larger the better," "the smaller the better," and "the specified target value is the best."

The objective of the parameter design is a matter of choosing a product or process condition that yields the best performance measure. In other words reducing the variation of response from the target while controlling the mean response toward the target is the ultimate goal of the parameter design.

These settings are determined by (1) systematically varying the settings of design parameters in the experiment and (2) comparing the effect of noise factors for each test run.

The parameter design achieves this goal by setting up an inner array and an outer array that constitute an orthogonal array such as $L_8(2^7)$ or $L_9(3^4)$. This array is assigned with control factor parameters, while the outer array is also an orthogonal array. This latter array is assigned with noise factor parameters.

TABLE 13 Parameter Design for Weight Watch Experiments

Outer Array			Row		1	2	3	4	
			Column						
			1	SEX	1	1	2	2	
			2	ETH	1	2	1	2	
			3		1	2	2	1	
Column	Inner Array 1	2			Observed WTLOSS				Performance Statistics
Row	DIET	EXR	3	4					Z_i
1	1	1	1	1	y_{11}	y_{12}	y_{13}	y_{14}	Z_1
2	1	2	2	2	y_{21}	y_{22}	y_{23}	y_{24}	Z_2
3	1	3	3	3	y_{31}	y_{32}	y_{33}	y_{34}	Z_3
4	2	1	2	3	y_{41}	y_{42}	y_{43}	y_{44}	Z_4
5	2	2	3	1	y_{51}	y_{52}	y_{53}	y_{54}	Z_5
6	2	3	1	2	y_{61}	y_{62}	y_{63}	y_{64}	Z_6
7	3	1	3	2	y_{71}	y_{72}	y_{73}	y_{74}	Z_7
8	3	2	1	1	y_{81}	y_{82}	y_{83}	y_{84}	Z_8
9	3	3	2	1	y_{91}	y_{92}	y_{93}	y_{94}	Z_9

8.3. Weight-Watch Experiments

Table 13 illustrates how the parameter design is set up for the weight-watch experiment.

Placed in the lower left corner is a $L_9(3^4)$. Columns 1 and 2 are assigned as DIET and EXERCISE, respectively, both having three levels. Placed in the upper right corner is a transposed outer array $L_4(2^3)$. In this outer array the transposed column 1 is assigned the noise factor SEX while column 2 is assigned another noise factor ETHNIC. The two arrays are arranged in such a manner that each of the nine control factor parameter combinations cross all of the four noise factor parameter combinations. Thus, a total of 36 observed WTLOSS pieces of date data are obtained and displayed as shown.

The subsequent analysis of data consists of computing performance statistics based on a formula such as

$$Z_i = (-10)\log\ 1/n \sum_{i=1}^{n} (1/y_i^2)$$

for a ''greater-the-better'' case like this.

The next steps follow the same flow of analysis using an ANOVA table as seen in the preceding sections.

9. THE STRATEGIES OF EXPERIMENTS (Du Pont Co. 1988)

The practice of experimentation is a matter of problem solving. It is also a learning process rendered through step-by-step development of know-how as well as know-why. However, it takes sound and smart strategies to reach this goal effectively and efficiently.

A complete set of strategies consists of three major components: strategy for screening designs, strategy for parameter designs, and strategy of response surface designs.

The main purpose of the strategy of screening designs is to screen out the significant factors from various possible factors selected for experiment.

However, in order to keep the cost of experimentation from being too high, two levels for each factor are used in general. Consequently, 2^k orthogonal arrays are most oftenly used in the screening designs.

The strategy of parameter designs is practiced after the strategy of screening designs. At this stage, only the factors that are found to be significant are included in the experiments. As the number of factors is reduced, the experiment is now focused at the finding and determination of the optimal condition. The 3^k orthogonal arrays are commonly used in this stage.

The strategy of response surface designs takes place after the parameter design strategy. All of the data collected in the repeatability experiments are then utilized in the database used in conjunction with the response surface designs.

The elaboration of response surface designs is beyond the scope of this article. For more treatment of the subject, see Cornell (1990), Khuri and Cornell (1987), and Montgomery (1991).

10. CONCLUSION

The success of experimentation relies on the following issues:

1. Thorough consideration of all technical, statistical, and administrative aspects
2. Sound planning of the experimentation
3. Effective implementation
4. Proper observation of good experimental practices
5. Smart adoption of strategies of experimentation

REFERENCES

Anderson, V. L., and McLean, R. A. (1974), *Design of Experiments,* Marcel Dekker, New York.

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

Bechhofer R. E., Santner, T. J., and Goldsman, D. M., *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons,* Wiley Series in Probability and Statistics, John Wiley & Sons. New York, 1995.

Bhote, K. R. (1991), *World Class Quality: Using Design of Experiments to Make It Happen,* AMACOM, New York.

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

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

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

Daniel, C. (1976), *Application of Statistics to Industrial Experiments,* John Wiley & Sons, New York.

Dean, A., and Voss, D. T. (1999), *Design and Analysis of Experiments,* Springer, New York.

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

Diamond, W. J. (1997), *Practical Experiment Designs for Engineers and Scientists,* John Wiley & Sons, New York.

Du Pont Co. (1988), *Strategy of Experimentation,* Du Pont Quality and Technology Center, Newark, NJ.

Gryna, F. M., and Juran, J. M. (1993), *Planning and Analysis of Quality: From Product Development Through Use,* 3rd Ed., McGraw-Hill, New York.

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

Hunter, J. S. (1998), "Design and Analysis of Experiments," in *Juran's Quality Handbook,* 5th Ed. J. M. Juran and A. B. Godfrey, Eds., McGraw-Hill, New York.

John, W. M. (1998), *Statistical Design and Analysis of Experiments,* Society for Industrial and Applied Mathematics, Philadelphia.

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

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

Montgomery, D. C. (1996), *Design and Analysis of Experiments,* 4th Ed., John Wiley & Sons, New York.

Peace, G. S. (1989), *Taguchi Methods: A Hands-on Approach,* Addison-Wesley, Reading, MA.

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

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

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

Taguchi, G., and Wu, Y. I. (1980), *Introduction to Off-line Quality Control,* Central Japan Quality Control Association, Nagoya, Japan.

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

Winer, B. J., Brown, D. R., and Michels, K. M. (1991), *Statistical Principles in Experimental Design,* McGraw-Hill, New York.

ADDITIONAL READINGS

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

Boniface, D. R., *Experiment Design and Statistical Methods: For Behavioral and Social Research,* CRC Press, Boca Raton, FL, 1995.

Cleveland, W. S., *Visualizing Data,* Hobart Press, Summit, NJ, 1993.

Cobb, G. W., *Introduction to Design and Analysis of Experiments,* Springer, New York.

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

Cornell, J. A., *How to Run Mixture Experiments for Product Quality,* Rev. Ed., American Society for Quality Press, Milwaukee, 1990.

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

Gad, S. C., *Statistics and Experimental Design for Toxicologists,* CRC Press, Boca Raton, FL, 1998.

Grove, D. M., and Davis, T. P., *Engineering Quality and Experimental Design,* John Wiley & Sons, New York, 1991.

Gunst, R. F., and Mason, R. L., *How to Construct Fractional Factorial Experiments,* American Society for Quality Press, Milwaukee, 1991.

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

Luftig, P. D., and Jeffrey, T., *Design of Experiments in Quality Engineering,* McGraw-Hill, New York, 1998.

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

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

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

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