DESIGN AND ANALYSIS OF SENSORY OPTIMIZATION

Maximo C. Gacula, Jr., Ph.D.
THE DIAL TECHNICAL AND ADMINISTRATIVE CENTER
SCOTTSDALE, ARIZONA

FOOD & NUTRITION PRESS
TRUMBULL, CONNECTICUT 06611 USA

DESIGN AND ANALYSIS
OF
SENSORY OPTIMIZATION

F
N
P

PUBLICATIONS IN FOOD SCIENCE AND NUTRITION

Books

VIEWPOINTS AND CONTROVERSIES IN SENSORY SCIENCE, H.R. Moskowitz *et al.*
PIONEERS IN FOOD SCIENCE, VOL. 2, J.J. Powers
DRY-CURED MEAT PRODUCTS, F. Toldrá
VEROCYTOTOXIGENIC *E. coli*, G. Duffy, P. Garvey and D.A. McDowell
OPEN DATING OF FOODS, T.P. Labuza and L.M. Szybist
NITRITE CURING OF MEAT: *N*-NITROSAMINE PROBLEM, R.B. Pegg and F. Shahidi
DICTIONARY OF FLAVORS, D.A. DeRovira
FOOD SAFETY: THE IMPLICATIONS OF CHANGE, J.J. Sheridan *et al.*
FOOD FOR HEALTH IN THE PACIFIC RIM, J.R. Whitaker *et al.*
DAIRY FOODS SAFETY: 1995-1996, A COMPENDIUM, E.H. Marth
OLIVE OIL, SECOND EDITION, A.K. Kiritsakis
MULTIVARIATE DATA ANALYSIS, G.B. Dijksterhuis
NUTRACEUTICALS: DESIGNER FOODS III, P.A. Lachance
DESCRIPTIVE SENSORY ANALYSIS IN PRACTICE, M.C. Gacula, Jr.
APPETITE FOR LIFE: AN AUTOBIOGRAPHY, S.A. Goldblith
HACCP: MICROBIOLOGICAL SAFETY OF MEAT, J.J. Sheridan *et al.*
OF MICROBES AND MOLECULES: FOOD TECHNOLOGY AT M.I.T., S.A. Goldblith
MEAT PRESERVATION, R.G. Cassens
PIONEERS IN FOOD SCIENCE, VOL. 1, S.A. Goldblith
FOOD CONCEPTS AND PRODUCTS: JUST-IN-TIME DEVELOPMENT, H.R. Moskowitz
MICROWAVE FOODS: NEW PRODUCT DEVELOPMENT, R.V. Decareau
DESIGN AND ANALYSIS OF SENSORY OPTIMIZATION, M.C. Gacula, Jr.
NUTRIENT ADDITIONS TO FOOD, J.C. Bauernfeind and P.A. Lachance
NITRITE-CURED MEAT, R.G. Cassens
CONTROLLED/MODIFIED ATMOSPHERE/VACUUM PACKAGING, A.L. Brody
NUTRITIONAL STATUS ASSESSMENT OF THE INDIVIDUAL, G.E. Livingston
QUALITY ASSURANCE OF FOODS, J.E. Stauffer
SCIENCE OF MEAT & MEAT PRODUCTS, 3RD ED., J.F. Price and B.S. Schweigert
NEW DIRECTIONS FOR PRODUCT TESTING OF FOODS, H.R. Moskowitz
PRODUCT DEVELOPMENT & DIETARY GUIDELINES, G.E. Livingston, *et al.*
SHELF-LIFE DATING OF FOODS, T.P. Labuza

Journals

JOURNAL OF FOOD LIPIDS, F. Shahidi
JOURNAL OF RAPID METHODS AND AUTOMATION IN MICROBIOLOGY, D.Y.C. Fung, M.C. Goldschmidt and D.H. Kang
JOURNAL OF MUSCLE FOODS, M.S. Brewer
JOURNAL OF SENSORY STUDIES, M.C. Gacula, Jr.
FOODSERVICE RESEARCH INTERNATIONAL, P.L. Bordi
JOURNAL OF FOOD BIOCHEMISTRY, N.F. Haard and B.K. Simpson
JOURNAL OF FOOD PROCESS ENGINEERING, D.R. Heldman and R.P. Singh
JOURNAL OF FOOD PROCESSING AND PRESERVATION, B.G. Swanson
JOURNAL OF FOOD QUALITY, J.J. Powers
JOURNAL OF FOOD SAFETY, T.J. Montville and K.R. Matthews
JOURNAL OF TEXTURE STUDIES, M.C. Bourne, T. van Vliet and V.N.M. Rao

Newsletter

FOOD, NUTRACEUTICALS AND NUTRITION, P.A. Lachance and M.C. Fisher

DESIGN AND ANALYSIS OF SENSORY OPTIMIZATION

Maximo C. Gacula, Jr., Ph.D.
THE DIAL TECHNICAL AND ADMINISTRATIVE CENTER
SCOTTSDALE, ARIZONA

FOOD & NUTRITION PRESS
TRUMBULL, CONNECTICUT 06611 USA

FOOD & NUTRITION PRESS, INC.

Trumbull, Connecticut 06611 USA

Library of Congress Catalog Card Number: 93-70222
ISBN: 0-917678-31-1

Printed in the United States of America

To my wife Amalia, for her understanding,
appreciation, and love

PREFACE

The ever-changing structure of research and development, changes in market demands, and the explosion of computing technology, resulted in the need of a book to aid professionals to survive these changes. Professionals now have ready access to personal computers equipped with spreadsheets and statistical software to create experimental designs and subsequent analysis of the data; this access results in a faster rate of product development. An important aspect during product development in the foods and consumer product industries is the role of sensory evaluation. An increasing number of companies requires that products should undergo extensive sensory tests before they are placed in the market. In several cases, products fail because of poor sensory properties, either on the basic tastes, i.e., sweetness, sourness, saltiness, bitterness, or in the case of consumer products, on skin-feel, fragrance, lather properties, etc.

In the last decade, a number of publications have been written to address the design and analysis of sensory and consumer studies. Unlike the previous publications, this book addresses two areas: product optimization incorporating Total Quality concepts and claims substantiation. Although optimization and Total Quality techniques have been known for some time, they have not been used extensively as their usefulness and importance would justify. It should be noted, however, that in the last few years several corporations in the United States and abroad have changed culture to embrace Total Quality. At present, claim substantiation is not well-addressed in sensory evaluation. The importance of sensory claims as seen in televisions, newspapers, and product brochures, justifies the need to increase focus on this area in sensory research.

Briefly, *Design and Analysis of Sensory Optimization*, provides sensory analysts, research professionals, statisticians, and students, the design and techniques of data analysis, from the inception of the project to evaluation of prototypes, their sensory optimization and selection of the final formula for consumer testing, to claims development for advertising purposes. The use of Taguchi's signal-to-noise ratio to select robust formulas in terms of consumer perception is discussed.

The level of this book requires completion of basic statistics. It was written for research professionals and others who can analyze their experimental data using a personal computer. This book can also be used as a text in applied statistics, as well as a reference material for industrial statistician working in consumer product industries.

In this book, several software packages were used in the examples to illustrate methods of data analysis:

SAS (Statistical Analysis System), a trademark of SAS Institute, Inc.

X-STAT, a trademark of Wiley Professional Software, John Wiley & Sons, Inc.

DESIGN-EXPERT, a trademark of Stat-Ease, Inc.

STATISTIX, a trademark of NH Analytical Software

The author wishes to thank several publishers and professional associations for permission to reproduce various materials used in this book. I am indebted to Dr. P. Dunn-Rankin for Appendix Table C.

I am deeply thankful to The Dial Corporation Technical and Administrative Center management for providing an excellent atmosphere for professional growth, and in particular to Dr. Richard J. Bertozzi.

Lastly, my appreciation to the Food & Nutrition Press for their unconditional support and assistance in the production of this book. My indebtness goes to Christine Murphy.

Any remaining error in this book is the responsibility of the author.

MAXIMO C. GACULA, JR.

CONTENTS

INTRODUCTION

In the last two decades, we have witnessed the specialization of statistical applications in various areas. Books have been published aimed at professionals in chemistry, biology, engineering, psychology, and many other scientific disciplines. Although the statistical assumptions remained intact for each application, some of them are not met. In particular, many properties of statistical methods are not satisfied in sensory data creating problems in the use of many statistical procedures. As a result, special attention is given to the choice of design and analysis of sensory data.

The wide availability of computing technology, both in hardware and software, resulted in an explosion of statistical use never before experienced. This brings statistics closer to the statistical users.

With this in view, the aim of this book is to bring together statistical applications and computing to the user in a manner that needs minimal knowledge of statistics and computers. This book deals with the application of statistics and computers to product optimization, experimental design, and analysis of sensory data.

1.1 STATISTICAL INFERENCE

In studying the behavior of a certain population, for example, the car buying habits of 20 to 30 years olds, one has to sample this population of interest, since it is impossible to study all members of the population. In order to obtain a representative sample of this population, random samples are obtained following an accepted random sampling procedure. Then these samples of individuals are studied based on pre-determined characteristics, such as color and type of the cars they purchased, price range, etc. Sample statistics are computed for each characteristic. Based on these statistics, inductive statements are made about the population being studied. The whole process is called statistical inference. Briefly, statistical inference consists of the following general steps:

1. Formulation and testing of hypothesis. This involves setting-up of the hypotheses, collection of data, choice of a test statistic, and specification of a decision rule for accepting or rejecting the null hypothesis.

2. Estimation of population parameters by computing sample statistics. Sample statistics would be an accurate description of the population parameters if they were obtained by an appropriate experimental design.

Regardless of how the experiment is conducted, sample statistics are subject to random variation. Random variation allows one to attach specified amount of confidence in making statistical inference on the results of experiments. Obviously, if the variation is large in the data it would be difficult to make a reliable inference. It is in the use of a sound experimental design that variation can be minimized if not controlled. Another tool to control variation, whether random or systematic, is the statistical analysis itself. These tools include covariance analysis, transformation, re-expression of data to a standard scale, use of averages to represent observations, use of nonparametric statistics, and others.

1.2 EXPERIMENTAL DESIGN

The key to a successful outcome in an experiment is the proper planning and execution of the work to be done. The planning and execution of the work belong to an area in statistics called experimental design. Briefly, experimental design is an organized approach to the collection of experimental data. This approach defines the population to be studied, the randomization process, the administration of treatments, the sample size requirement, and the method of statistical analysis. The rewards of a well-designed study are the reduction of cost of experimentation, the ease of interpretation of results, and the procurement of good data to result in a useful and meaningful outcome.

Depending on the reader's formal training and experience, there are books that deal with experimental design and statistical analysis (Cochran and Cox 1957; Federer 1955; Steel and Torrie 1960; Kempthorne 1952; Kirk 1968). For applications in the food and consumer research useful books are by Amerine *et al.* (1965), Amerine and Roessler (1976), Gacula and Singh (1984), Stone and Sidel (1985), O'Mahony (1986), Piggott (1986), Meilgaard *et al.* (1987), and most recently by Moskowitz (1988).

1.3 SAMPLE SIZE

The number of observations plays a major role in the design of experiments. For example, in the planning of a consumer test, one of the key questions sensory analysts often ask is the number of panelists to be recruited to provide sufficient sensitivity for detecting a difference at the specified significance level. It is important to briefly introduce some statistical concepts and terminologies essential for understanding the estimation of sample size.

In a consumer test, the sensory analyst often desires to determine whether two formulations differ with respect to their mean scores on prescribed sensory attributes. The statistical formulation of this statement is in terms of the null hypothesis denoted by H_0. That is, H_0: $\mu_1 - \mu_2 = 0$. In other words the null hypothesis states that there is no difference between the two population means μ_1 and μ_2, other than that due to random variation. If H_0 is not true we reject it, and accept the so-called alternative hypothesis, H_a: $\mu_1 - \mu_2 \neq 0$. The estimation of μ_1 and μ_2 from random samples is always subject to uncertainty. Therefore, it is possible to wrongly reject H_0 and commit a Type I error; the probability of this error being committed is denoted by α. On the other hand, it is also possible to wrongly accept H_0 and commit a Type II error, the probability of which is denoted by β. Table 1.3-1 shows the two types of errors. The quantity $1 - \alpha$ is known as the confidence level of the test and $1 - \beta$ is called the power of the test. The power of the test denotes the ability of the test statistic to reject a false hypothesis. See Chapter 9 for further discussion.

There is another point in hypothesis testing that should be mentioned, which is the value of the null hypothesis. It is not necessarily equal to zero; any appropriate values or equality of parameters can be used in relation to the study. For example, the following null hypotheses can be tested with an appropriate experimental design:

$$H_o: \mu_1 = \mu_2 = \mu_3 = 0$$

$$H_o: \mu_1 - \mu_2 > 1 \text{ log}$$

$$H_o: \mu_1 - \mu_2 = 5.5$$

The method of data analysis for testing the above hypotheses follows the experimental design.

In determining the sample size N, we must consider the Type I and Type II errors to obtain a sensitive test. A general formula is

$$N = [(Z_\alpha + Z_\beta)^2\sigma^2]/(\mu_1 - \mu_2)^2 \qquad (1.3\text{-}1)$$

where Z is the standard deviation of the standard normal distribution, σ^2 is the variance, and $\mu_1 - \mu_2$ is the desired difference to be detected; as given by (1.3-1), N becomes larger as this difference gets smaller and vice versa. The value for σ^2 may be obtained from historical data, and a value of $\sigma^2 = 0.5$ to 1.0 seems appropriate for rating scales between 5 and 9 categories. Selected values of α and β useful in sample size calculation are shown in Table 1.3-2.

For example, let us determine the number of panelists needed so that one can detect a difference of 0.5 on the 9-point scale with a significance level of $\alpha = 0.05$ and a power of 0.90. Assuming that $\sigma^2 = 1.0$, one obtains

$$N = [(1.960 + 1.645)^2 1^2]/0.5^2 = 52$$

Table 1.3-1
Two types of errors in hypothesis testing.

Decision	If H_o is:	
	True	False
Accept H_o	No error $(1 - \alpha)$	Type II error β
Reject H_o	Type I error α	No error $(1 - \beta)$

Note: α = Probability of rejecting the null hypothesis when it is true.
β = Probability of accepting the null hypothesis when it is false.

panelists per treatment. In practice, one should recruit more than 52 panelists to compensate for "dropouts."

1.4 RANDOMIZATION

Fisher (1960) in his classic book, *The Design of Experiments*, stated that randomization is the physical basis of the significance test. This statement is a fundamental point in the use of randomization in experimental design—to guarantee that the statistical test will have a valid significance level. In general, randomization is defined as the random allocation of experimental materials to treatments so that each material has an equal chance of being assigned to one of the treatments. In this definition, treatment refers to conditions or a combination of conditions from which the experimental outcome is to be observed. In sensory/consumer testing, a common form of randomization is the random distribution of panelists to a specified group. By doing this, the uncontrolled variation among panelists is distributed to treatment groups, and the treatment effect is therefore similarly affected, resulting in cancellation of the effect of variation. Another form is the random ordering of sample presentation. In a two-product test, systematic randomization is often used for simplicity; i.e., the order of sample presentation is alternating, for example, AB, BA, AB, and so on. However, the danger of this method is that the pattern of systematic bias may follow the alternating sample order to result in a bias estimate of treatment effects. This is dangerous particularly in clinical irritation testing, because the judge's evaluation on the subject's treated skin is done at one sitting; judges may follow the AB-BA pattern during evaluation. Furthermore, clinical irritation test is generally characterized by a small panel size, and this pattern is likely to occur compared to that of larger panel size, such as in consumer testing. When the order of evaluation is randomized, the systematic bias becomes a random variation.

Although the goal of randomization is the same for all experimental designs, to isolate and/or distribute random variation in the data, each randomization procedure varies with the experimental design.

Table 1.3-2
Selected values of Z_α and Z_β for sample size calculation.

α	Z_α	β	Z_β
0.01	2.576	0.10	1.645
0.05	1.960	0.20	1.282
0.10	1.645	0.30	1.036

Example 1.4-1

Let us illustrate the effect of randomization on statistical inference using actual experimental data. It is shown in this illustration that without the use of randomization in the design, a wrong conclusion is bound to happen. In a clinical study two identical products, denoted by A and B, were compared on the basis of deodorant efficacy using 16 panelists. The assignment of products to the right and left armpits was randomized. The randomization gave the following assigments:

Panelist:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Right:	B	B	B	A	A	A	A	B	A	A	A	B	B	B	A	B
Left:	A	A	A	B	B	B	B	A	B	B	B	A	A	A	B	A

The judges were instructed to always sniff the right axilla first followed by the left axilla.

Since products A and B are identical, the null hypothesis is H_0: $\mu_A - \mu_B = 0$. If we disregard the randomization schedule in the analysis, we compute the difference D between right and left, that is D = Right − Left, instead of D = A − B. The result of both calculations is given in Table 1.4-1 for degrees of malodor using a 7-point intensity scale. Statistically significant comparisons were obtained by the Right − Left calculations in days 1 and 3, which are obviously incorrect since products A and B were identical. In days 1, 3, and 4, the right axilla—the first to be sniffed, was perceived to have more malodor than the left axilla. To illustrate, consider the data for the first four panelists on day 1:

Panelist	1		2		3		4	
Randomization order	B	A	B	A	B	A	A	B
Score	0	0	0	0	1	0	1	0
Right − Left	0		0		1		1	
A − B	0		0		−1		1	

The right axilla, the first to be sniffed, for panelists 3 and 4 were rated higher than that of the left axilla. This results in a mean difference of (0 + 0 + 1 + 1)/4 = 0.5 for the Right − Left calculations. This difference is an example of the so-called

Table 1.4-1
Test of hypothesis comparing two identical products with (A − B) and without (Right − Left) considerations of the randomization schedule.

Comparison	Day	Mean difference	Std. dev.	Sig. prob.
A − B	1	−0.033	0.129	0.334
Right − Left		0.156	0.239	0.020*
A − B	2	−0.067	0.372	0.499
Right − Left		−0.094	0.272	0.188
A − B	3	−0.033	0.297	0.670
Right − Left		0.219	0.256	0.004*
A − B	4	0.033	0.481	0.792
Right − Left		0.125	0.465	0.300

Note: Std. dev. = standard deviation of the difference.
Sig. prob. = significance probability.
* Significant difference.

position/order bias. The mean difference for the A − B calculation is (0 + 0 − 1 + 1)/4 = 0.0, a result consistent with the null hypothesis.

For all the evaluation days, no significant differences were observed for the A − B comparisons, in which the calculation was based in accordance with the randomization schedule. This result strongly demonstrates the physical validity of randomization in sensory experiments.

1.5 ANALYSIS OF VARIANCE

The analysis of variance (ANOVA) is a well-known technique for partitioning the total variation of observations into component parts to facilitate the testing of various hypotheses of interest. The sources of variation in the analysis of variance depend on the experimental design used in the collection of the data. Suppose that one is investigating four treatments and the response variable is a chemical analysis of water activity in certain food product. Samples from each treatment were secured and analyzed for water activity. The first step in the analysis is to write the statistical model for each observation as follows:

$$X_{ij} = \mu + T_i + E_{ij} \qquad (1.5\text{-}1)$$
$$i = 1, 2, \ldots, n$$
$$j = 1, 2, \ldots, k$$

where X_{ij} the observed water activity for the ith treatment and jth observation, μ the grand mean for all observations, T_i the effect of the ith treatment with constraint

Table 1.5-1
One-way analysis of variance

Source of Variance	DF	SS	MS	F-ratio
Total	N − 1	SSTO		
Between treatments	k − 1	SSA	MSA	MSA/MSE
Error	N − k	SSE	MSE	

Note: N = Total number of observations for the entire experiment = n(k).
k = Number of treatments or groups.

that $\Sigma T_i = 0$, and E_{ij} are random errors assumed to be normally and independently distributed with variance σ_e^2.

It is seen in Eq. (1.5-1) that treatment is the only main source of variation. The form of the analysis of variance for this model is shown in Table 1.5-1. When there is one main source of variation, we refer to it as one-way ANOVA. Obviously, when there is more than one main source, we refer to them as two-way, three-way, and so on.

The main sources of variation are known as main effects and the dependencies between main sources are called interaction effects. In this table and the succeeding ANOVA tables to be presented, the degrees of freedom are denoted by DF, the sums of squares by SS, and the mean squares by MS. Note that MS is obtained by SS/DF. As shown in Table 1.5-1 for one-way ANOVA, the DF is found by subtracting 1 from the total number of observations for the case of total DF, subtracting 1 from the number of treatments for the treatment DF, and the error DF obtained by difference or N − k. The same principle is used to find DF for other ANOVA tables. To determine whether the four treatment effects are significantly different from each other, the F-ratio statistic is computed, which is the ratio between the mean square of interest and the error mean square. For the ANOVA in Table 1.5-1, the F-ratio is F = MSA/MSE. If the F ratio is greater than the tabled F value (Table A, Appendix), one concludes that at least one treatment mean is significantly different from one of them at the specified significance level.

1.6 MULTIPLE COMPARISON TESTS

The F ratio statistic, when significant, indicates that at least one treatment mean is significantly different from one or more treatment means in the study. To obtain a pair-wise comparison of all treatments, one of the many multiple comparison procedures is conducted. These procedures are given in several publications such as Steel and Torrie (1960), Kirk (1968), Gacula and Singh (1984), and O'Mahony (1986). The application of the Duncan's multiple range test (Duncan 1955) and a rank sum multiple comparison test (Dunn-Rankin 1965; Nemenyi 1963) is given in this section.

Duncan's Multiple Range Test

Let p denote the number of means in a subset of k means. For p means, the critical difference of the test known as the shortest significant range denoted by R_p is

$$R_p = r_{\alpha,p,DF}\,(\sqrt{MSE/n}) \tag{1.6-1}$$

where $r_{\alpha,p,DF}$ is tabulated in Table B for $\alpha = 0.05$, 0.01 and for several values of p and error DF. To determine whether any two means are significantly different, the means are first arranged in order of magnitude. Then pairwise mean differences d_{ij} and their corresponding R_p are computed. A significant difference is declared when $d_{ij} > R_p$.

For example assume that the means, in order of magnitudes, are $X_1 > X_3 > X_2$. Then the pairwise differences and their corresponding subset size p are as follows:

d_{ij}	p
$d_{12} = X_1 - X_2$	3
$d_{13} = X_1 - X_3$	2
$d_{32} = X_3 - X_2$	2

For difference d_{12}, use $p = 3$ for finding $r_{\alpha,p,DF}$ in Table B since there are three means involved in the range; for d_{13} use $p = 2$, and so forth. A numerical example of the use of Duncan's multiple range test is given by Example 3.1-1.

Rank Sum Multiple Comparison Test

Ranking of items or treatments is a common sensory method for gathering data. There are various methods (Dunn 1964; Steel 1961; Tobach *et al.* 1967; Wilcoxon and Wilcox 1964) for the statistical analysis of rank data that vary in complexity from one another. One method is the conversion of ranks into normal scores prior to the application of the analysis of variance as illustrated by Larmond (1967), Amerine and Roessler (1976), and Gacula and Singh (1984). The Steel procedure is another method that is used after the observations have been assigned ranks according to their magnitudes; this procedure is tedious because of the need for re-ranking the data for succeeding comparisons, and the statistical tables for test of significance is limited for sensory work. Another method is by Kramer (1963) and Kramer *et al.* (1974), but it is not recommended because of a questionable method used in the calculation of the statistical tables and the ensuing problems of interpretation of experimental results (Joanes 1985).

The Nemenyi (1963) and the Dunn-Rankin (1965) multiple comparison test, which is based on the distribution of range of rank totals of treatments, is considered here. A statistical table of this test for obtaining significance between pairwise comparison has been extended by Dunn-Rankin to include 15 treatments and 500 panelists. Table

C in the appendix is an abbreviated version of this table. The experimental setup for the use of this procedure consists of N panelists ranking k treatments from 1 through k. The steps are summarized as follows:

1. Calculate the total of ranks for each treatment and arrange them in descending order of magnitude.
2. Calculate the pairwise differences, d_{ij}, between rank totals R_i. For k treatments, the total number of pairwise differences is $k(k-1)/2$.

$$d_{ij} = R_i - R_j, \; i > j \tag{1.6-2}$$

3. Consult Table C and declare a significant difference at the α level, if d_{ij} is greater than the tabled critical value at the indicated number of treatments and number of panelists.

Note that in the first step, if the observed data are not in the form of ranks, they are assigned ranks according to their magnitude. Numerical examples of a rank sum multiple comparison test are given in Chapters 3 and 7.

1.7 SOME USEFUL TOOLS FOR DATA ANALYSIS

Results from the statistical analysis of data are influenced by the method used in the analysis. Therefore, it is very important to have an experimental design before the experiment is to begin. The method of analysis depends on the experimental design used in the collection of the data. When the analysis is completed, the effects of the variables in the model are examined and tested for statistical significance. This test provides information on whether the effects are real and not due to sampling errors. The effects, not due to treatments, that are large may be removed through adjustment of data to provide a better estimate of treatment effects that is free from extraneous variation.

In sensory evaluation work it is well-known that the panelist is a major source of variability in the data. See for example, the recent work by Finkey *et al.* (1988), which showed that panelist accounted for over 80% of the total variability of irritation scores. This variability is mostly the result of the differences on how judges used the length of the scale, threshold differences among judges with respect to certain stimulus, and most of all the differences in panelist response to skin irritation stimuli. These sources of variability may not be important in paired comparison analysis, as the statistical evaluation is based on differences (See Chapter 2). Working with differences eliminates this source. However, when one is not working with differences, a problem may arise especially when the data are subjected to correlation/regression analysis. To minimize this problem, the data may be re-expressed in another scale to remove variabilities due to threshold, scale usage, and differences due to panelist response.

Deviation From The Mean

The re-expression of data by deviation from the mean forces the scale origin to zero within a particular subset (Gacula *et al.* 1971). A judge or panelist is an example of a subset, and an average estimated from them is a subset mean. The choice of subset mean is critical in the use of this method of analysis because a wrong choice may result in the removal of other effects in the model. Consider the simple model, Eq. (1.7-1):

$$Y_{ij} = \mu + T_i + P_j + E_{ij} \qquad (1.7\text{-}1)$$
$$i = 1, 2, .., k$$
$$j = 1, 2, .., n$$

where Y_{ij} are observed scores of the ith treatment and jth panelist; μ the grand mean across treatments and panelists; T_i the effect of the ith treatment; P_j the effect of the jth panelist; and E_{ij} are random errors not accounted for by the model. Table 1.7-1 shows the layout of the data for model (Eq. 1.7-1). The re-expressed data D_{ij} is obtained by subtracting each observation from its corresponding subset mean. Thus we have the following equations:

$$\begin{aligned}
D_{11} &= Y_{11} - M_{.1} \\
D_{21} &= Y_{21} - M_{.1} \\
D_{31} &= Y_{31} - M_{.1} \\
D_{12} &= Y_{12} - M_{.2} \\
D_{22} &= Y_{22} - M_{.2} \\
D_{23} &= Y_{23} - M_{.2} \\
&\bullet \quad \bullet \quad \bullet \\
&\bullet \quad \bullet \quad \bullet \\
&\bullet \quad \bullet \quad \bullet \\
D_{kn} &= Y_{kn} - M_{.n}
\end{aligned} \qquad (1.7\text{-}2)$$

The resulting D_{kn}s are the raw data to be used in the statistical analysis. Note that for each panelist, the subset mean for the re-expressed data is equal to zero because each observation was expressed as a deviation from the subset mean. The use of subset mean as a pivot is similar to blocking the effects of panelist in randomized complete block design (See Chapter 3). Thus the comparison of treatments is made on within panelist basis which is more homogenous than the comparison across panelists.

Table 1.7-1
Layout of data for model (Eq. 1.7-1).

Panelist	Treatment	Observation	Subset mean
1	1	Y_{11}	
	2	Y_{21}	
	3	Y_{31}	$M_{.1}$
2	1	Y_{12}	
	2	Y_{22}	
	3	Y_{32}	$M_{.2}$
...	...	...	...
...	...	...	...
...	...	...	...
n	k	Y_{kn}	$M_{.n}$

Note: Subscript substituted by a "." indicates summation of values over the subscript.

Example 1.7-1

Let us illustrate the re-expression of data by the deviation from the subset mean using a portion of the data from an experiment correlating surfactant irritation obtained subjectively with fiber optic spectroscopy (Crowe *et al.* 1988). Three solutions of a surfactant, 0.5%, 1.0%, and 2.0% in deionized water were occluded on the right forearms of the panelist. After 24 h, the sites on the forearms were evaluated for erythema by two judges using a 5-point rating scale where 0 = no redness and 4 = extremely red. Table 1.7-2 shows the data for four panelists. The erythema scores in this table are the average of two judges. The Y values in the table are actual measurements and the D values are the re-expressed data using formula, Eq. (1.7-2). To obtain the re-expressed data, the mean for each panelist was computed across treatments. For panelist 1 the computations for the instrumental data are as follows:

$$M_{.1} = 0.195$$

$$D_{11} = 0.183 - 0.195 = -0.012$$

$$D_{21} = 0.198 - 0.195 = 0.003$$

$$D_{31} = 0.203 - 0.195 = 0.008$$

For erythema scores the result is

$$M_{.1} = 1.0$$

$$D_{11} = -1.0$$

$$D_{21} = 0.0$$

$$D_{31} = 1.0$$

Table 1.7-2
Instrumental and erythema data for Example 1.7-1.

Panelist	Treatment	Instrumental		Score	
		Y	D	Y	D
1	1	.183	−.012	0	−1.0
	2	.198	.003	1	0.0
	3	.203	.008	2	1.0
2	2	.208	−.002	1.5	−0.170
	3	.201	−.009	2.5	0.833
	1	.221	.011	1.0	−0.670
3	2	.203	.003	3	0.0
	3	.218	.018	4	1.0
	1	.178	−.022	2	−1.0
4	3	.183	.002	1.5	0.50
	1	.188	.007	1.0	0.0
	2	.173	−.008	0.5	−0.50

Note: Treatment 1 = 0.5%.
Treatment 2 = 1.0%.
Treatment 3 = 2.0%.
Application of treatments to sites on the forearms was done in balanced order.
Y = actual observed data.
D = re-expressed data.

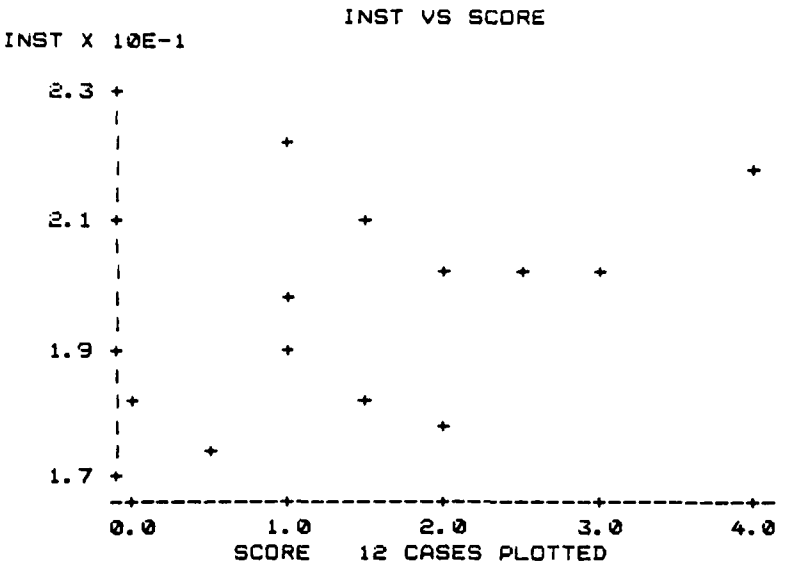


Fig. 1.7-1
Plot of actual data for Example 1.7-1.

and so on to the other panelists.

Figures 1.7-1 and 1.7-2 show the plot of the data shown in Table 1.7-2 for the actual and re-expressed data, respectively, to demonstrate the effectiveness of the technique for removing panelist effect. In this plot, the vertical axis is erythema scores and the horizontal axis the instrumental readings. The scatter of points about the regression line for the actual data are wider than those of the re-expressed data. The squared correlation coefficient for these data was also higher in the re-expressed data: 0.33 vs. 0.26 (Tables 1.7-3 and 1.7-4). Notice that the regression equation for the re-expressed data has no intercept (constant) because $\Sigma D_{ij} = 0$, the result of data re-expression. There was also higher significance probability for testing the slope of the regression line in the re-expressed data, $p = 0.0397$ vs. $p = 0.0886$, indicating lower variability in these data.

Plotting of Data

The oldest form of data analysis is plotting or graphing. We have illustrated this in Example 1.7-1 to study the relationship. Data can be depicted in the form of histograms, charts, and plots for two or more variables. For sensory data, histograms are very informative because they reveal bimodality of the distribution of judgments; they also reveal how the panelist used the length of the scale.

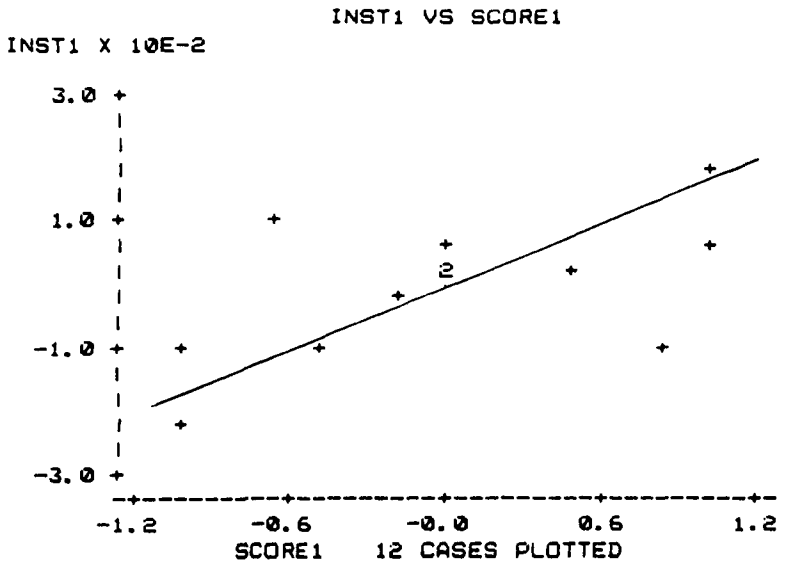


Fig. 1.7-2
Plot of re-expressed data for Example 1.7-1.

Table 1.7-3
Output from STATISTIX software for Example 1.7-1. SCORE AND INST are actual measurements.

UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF SCORE

PREDICTOR VARIABLES	COEFFICIENT	STD ERROR	STUDENT'S T	P
CONSTANT	-5.5820	3.8542	-1.45	0.1782
INST	36.905	19.567	1.89	0.0886

CASES INCLUDED	12	MISSING CASES	0
DEGREES OF FREEDOM	10		
OVERALL F	3.557	P VALUE	0.0681
ADJUSTED R SQUARED	0.1886		
R SQUARED	0.2624		
RESID. MEAN SQUARE	1.008		

In curve fitting analysis, it is first recommended to look at the plots of two or more variables before attempting to fit the data to a certain model. The scatterplot reveals possible form of the relationship between variables, thus saving time and effort in statistical modeling or fitting. As a diagnostic tool the scatterplot guides the data analyst on whether to use a transformation on the data, or to use a first or second degree equation for fitting. Attempts to fit the data to a model and solely use the R^2 criterion as a measure of goodness of fit without looking at the scatterplot could be misleading in the presence of outlying observations.

When a scatterplot is curvilinear some form of transformation is necessary before a useful correlation coefficient can be obtained, because a simple correlation coefficient defines a linear relationship. In a regression problem, when the relationship is linearized, only two parameters are estimated as opposed to nonlinear with more than two parameters to be estimated. The larger the number of parameters to be estimated from the data the more chances of introducing errors to the estimated value of the dependent variable.

Plotting is a useful tool in interpreting interactions in analysis of variance. Interactions may be significant but can possibly be due to sampling errors. Thus the interpretation of interaction should be carefully done. Contour plotting and its usefulness in interpreting results from response surface analysis are discussed in Chapter 8.

Rejection of Outlying Observations

In the evaluation of data one often encounters unusual observations in the sense that they are extremely below or above expectations. These observations are often called outlier, wild, or aberrant. If they are truely outliers they should be discarded, as inclusion of these data points will affect the outcome of the statistical analysis, hence the conclusion of the experiment. In some instances, they are known in advance to be an outlying observation, such as the result of instrument malfunction, instrument miscalibration, or solution contamination. In this case one may discard the

Table 1.7-4
Output from STATISTIX software for Example 1.7-1. SCORE1 and INST1 are re-expressed data.

UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF SCORE1

NOTE: MODEL FORCED THROUGH ORIGIN

PREDICTOR VARIABLES	COEFFICIENT	STD ERROR	STUDENT'S T	P
INST1	37.195	15.943	2.33	0.0397

CASES INCLUDED	12	MISSING CASES	0
DEGREES OF FREEDOM	11		
OVERALL F	5.442	P VALUE	0.0397
ADJUSTED R SQUARED	0.2702		
R SQUARED	0.3310		
RESID. MEAN SQUARE	3.449E-01		

data with no further statistical test of significance. In another case, the data is suspiciously large or small with no apparent reasons, and it is advisable to conduct a statistical analysis to determine whether to reject or retain the suspicious data.

It is important to point out that subjective data obtained from hedonic, preference, or acceptability measurement do not contain outlying observations because this type of data is an expression of likes and dislikes, and therefore is never wrong. However, subjective data from intensity evaluation is subject to aberrant or wild judgment and can be discarded by a statistical test.

TEST PROCEDURES

In this section, three test procedures are discussed, each procedure being strictly applicable to specific situations.

The r Ratio Test

This test developed by Dixon (1950, 1951, 1953) computes the ratio of the interval between the aberrant and adjacent data to the total range on the assumption that the data follows the normal distribution. The critical values of the test for numbers of observations ranging from n = 3 to n = 25 are given in Table E. The steps in performing the procedure are as follows:

1. If the smallest observation is the suspected outlier, rank the n observations from smallest to largest, i.e.,

$$X_1 \leq X_2 \leq \ldots \leq X_n$$

2. If the largest observation is the suspected outlier, rank the n observations from largest to the smallest.

3. Compute the r ratio depending on the sample size n:

$$r_{10} = (X_2 - X_1)/(X_n - X_1), \quad \text{for } n = 3 \text{ to } 7$$

$$r_{11} = (X_2 - X_1)/(X_{n-1} - X_1), \quad \text{for } n = 8 \text{ to } 10$$

$$r_{21} = (X_3 - X_1)/(X_{n-1} - X_1), \quad \text{for } n = 11 \text{ to } 13$$

$$r_{22} = (X_3 - X_1)/(X_{n-2} - X_1), \quad \text{for } n = 14 \text{ to } 30$$

4. Reject the suspected observation at the specified level of significance if the computed ratio exceeds the critical value given in Table E in the Appendix.

Example 1.7-2. Consider the following scores using the 7-point off-flavor intensity scale from 10 trained panelists:

Panelist	Score
1	3
2	4
3	1
4	3
5	4
6	3
7	2
8	3
9	2
10	7

Here, $n = 10$, $\overline{X} = 3.20$, and $S = 1.62$. It is suspected that the score 7 given by panelist 10 is unusually large. By Step 1, the data are arranged in descending order as follows:

$$\underset{X_1}{7}, \quad \underset{X_2}{4}, \quad 4, \quad 3, \quad 3, \quad 3, \quad 3, \quad 2, \quad 2, \quad \underset{X_n}{1}$$

With $n = 10$, the r ratio is

$$r_{11} = (4 - 7)/(1 - 7) = 0.50$$

which exceeds the tabled value of 0.477 (Table E) at the 0.05 level, thus the score 7 is an outlier.

For the purpose of illustration, assume that the smallest observation is a suspect. Then the data are arranged in the ascending order of magnitude as follows:

$$1, \quad 2, \quad 2, \quad 3, \quad 3, \quad 3, \quad 3, \quad 3, \quad 4, \quad 4, \quad 7$$

Then

$$r_{11} = (2 - 1)/(7 - 1) = 0.17$$

which is obviously not statistically significant and thus the score 1 is not an outlier.

Note that rejection of outlying observation can be applied not only to individual observations but also to estimated mean values from various experiments. For example, estimates of shelf-life of a product conducted at different times of the year may contain an outlying estimate which is atypical of the true shelf-life of the product under normal condition of use. Here are some estimates of average shelf-life (days) of a product over a 4-year period illustrating extreme variability.

1978	35, 50
1979	50, 59
1980	80, 52
1981	60, 52

Are 35 and 80 days aberrant estimates? If the lowest value is included it may underestimate the true shelf-life and results in monetary loss to the producer. On the other hand, inclusion of the largest value may cause product spoilage on the shelves to result in expensive product recall.

The Grubbs Ratio Test

Grubbs (1950) and Grubbs and Beck (1972) proposed this method for the rejection of the largest or the smallest observation in a sample of size n from a normally distributed population. In addition, two largest observations or two smallest observations can be tested whether they are outliers. This method, to be denoted as the G ratio test, is based on the distribution of the ratio of sums of squares (SS) where the numerator of the ratio is the SS without the suspected observation and the denominator is the SS when all observations are included.

Table 1.7-5 shows the SS formulas for computing the various ratios for test of significance. Each test ratio is applied one at a time with a corresponding reduction in n, if an observation is declared an outlier in the first test. The critical values of the test are given in Tables F and G for levels of significance 0.01 to 0.10. The computed ratio must be equal to or less than the critical value G at the indicated significance level to declare the observation an outlier.

For testing significance of the largest observation the ratio

$$G_n = S_n^2/S^2$$

is computed where S_n^2 is the SS without the largest observation and S^2 is the total SS using all observations. Following similar reasoning, the statistic

Table 1.7-5
Formulas for computing sums of squares for the Grubbs Ratio test.

$S^2 = \Sigma(X_i - \bar{X})^2$, where $\bar{X} = \Sigma X_i/n$, $i = 1, 2, .., n$
$S_n^2 = \Sigma(X_i - \bar{X}_n)^2$, where $\bar{X}_n = \Sigma X_i/n - 1$, $i = 1, 2, .., n - 1$
$S_1^2 = \Sigma(X_i - \bar{X}_1)^2$, where $\bar{X}_1 = \Sigma X_i/n - 1$, $i = 2, 3, .., n$
$S_{n-1,n}^2 = \Sigma(X_i - \bar{X}_{n-1,n})^2$, where $\bar{X}_{n-1,n} = \Sigma X_i/n - 2$, $i = 1, 2, .., n - 2$
$S_{1,2}^2 = \Sigma(X_i - \bar{X}_{1,2})^2$, where $\bar{X}_{1,2} = \Sigma X_i/n - 2$, $i = 3, 4, .., n$

Note: Σ = summation sign, i.e., sum over the ith observation.

$$G_1 = S_1^2/S^2$$

is computed for testing the smallest observation. Use Table F for this test.

For testing whether the two largest observations are too large the ratio is

$$G_{n-1} = S_{n-1,n}^2/S^2$$

and the ratio

$$G_{1,2} = S_{1,2}^2/S^2$$

is for testing the two smallest observations. Use Table G for this test. The computed ratio must be greater than the tabled value to declare statistical significance.

Example 1.7-3

The days to failure based on a mean score of 3.5 (7-point intensity scale) for a particular packaged food product arranged in the ascending order are as follows: 49, 56, 59, 69, 70, 71, 73, 75, 76, 78, 101, 110. Determine whether the smallest, the the largest, or the two largest observations are aberrant.

Using the formulas in Table 1.7-5, the following sums of squares are obtained:

$$S^2 = 3270.96, \qquad S_1^2 = 2593.61, \qquad S_n^2 = 1850.55$$
$$S_{1,2}^2 = 2145.60, \qquad S_{n-1,n}^2 = 836.40$$

If the smallest observation is suspected to be aberrant, we compute the ratio

$$G_1 = 2593.61/3270.96 = 0.7929$$

which is not significant even at the 10% level since G1 > G at n = 12 (Table F). Thus this observation is not an outlier. Table 1.7-6 shows the calculations.

If the largest observation is suspect, the ratio is

$$G_n = 1850.55/3270.96 = 0.5658$$

which is also not significant since 0.5658 is greater than the tabled value of 0.5483.

Note that if the smallest observation was found to be an outlier, and we desire to test also for the largest observation, the sample size must be reduced by 1 in computing S^2. The calculations are illustrated in Table 1.7-7. Note that the calculation of S_n^2 is now based on n = 10. It is emphasized that a legitimate reason should exist that the observation is an outlier, otherwise this statistical tool will be abused resulting in obtaining nonrepresentative data.

To continue with the example, if the two smallest observations are suspect, the ratio is

$$G_{1,2} = 2145.60/3270.96 = 0.6560$$

which is far from significance even at the 10% level (Table G), thus both observations are not outliers.

To test for the two largest observations, the ratio is

$$G_{n-1,n} = 836.40/3270.96 = 0.2557$$

which is significant at the 5% level (between 2.5 and 5.0%, n = 12, Table G). Hence the observations 101 and 110 are outliers.

The u Test

The best method in the sense of power of the test for detecting outliers is the so-called u test as discussed by Dixon (1950, 1953). However, this test method depends on a past estimate of standard deviation and cannot be used when this estimate is not available. The method is used to test the smallest or the largest observation from a normally distributed population.

If the smallest observation X_1 is the suspected value, we compute the ratio

$$u_1 = (\overline{X} - X_1)/S$$

or compute

$$u_n = (X_n - \overline{X})/S$$

Table 1.7-6
Calculations for rejection of smallest observation by the Grubbs ratio test (Example 1.7-3).

Sample	Observation	$(X_i - \overline{X})^2$	$(X_i - \overline{X}_1)^2$
1	49	621.01	...
2	56	321.13	407.23
3	59	222.61	295.15
4	69	24.21	51.55
5	70	15.37	38.19
6	71	8.53	26.83
7	73	0.85	10.11
8	75	1.17	1.39
9	76	4.33	0.03
10	78	16.65	3.31
11	101	733.33	616.03
12	110	1301.77	1143.79
Sum	887	3270.96	2593.61
Mean $\overline{X}$	73.92		
Mean $\overline{X}_1$	76.18 (excluding observation 49)		

Note: $G_1 = 2593.61/3279.96 = 0.7929$, not significant, $G_1 > G$ (Table F).

if the largest observation is the suspected value. Here, S is the past estimate of standard deviation, X_1 the smallest observation, X_n the largest observation, and $\overline{X}$ the mean of n observations. The u test is particularly useful in analytical analysis/laboratory instrumentation work where n is relatively small. There is a limited use of this method in sensory evaluation because of the difficulty in obtaining and maintaining an accurate estimate of past standard deviation.

Appendix Table H gives the critical values of the test reproduced from Nair (1952). The critical values should be exceeded by the computed value of u to declare significance or rejection of the suspected observation.

Example 1.7-4

Consider the shelf-life data obtained over a 4-year period stated earlier where $\overline{X}_1 = 35$, $X_n = 80$, and $\overline{X} = 54.75$ days based on $n = 8$ estimates. Past data based on $v = 10$ estimates showed that $S = 12$. Determine whether the extreme estimates are aberrant. The calculations are as follows:

$$u_1 = (54.75 - 35)/12 = 1.65$$

$$u_n = (80 - 54.75)/12 = 2.10$$

At the 1% significance level, $u(n=8, v=10) = 3.59$ (Table H) which is not exceeded by the computed u statistic, hence both extremes are not outliers.

There is a simple procedure that uses the maximum Z score to detect outliers. Outliers are sometimes defined as those values for which the absolute of Z exceeds

Table 1.7-7
Calculations for rejection of largest observation by the Grubbs ratio test when the smallest observation was found to be an outlier (Example 1.7-3).

Sample	Observation	$(X_i - \overline{X})^2$	$(X_i - \overline{X}_n)^2$
1	...	...	...
2	56	407.23	282.24
3	59	295.15	190.44
4	69	51.55	14.44
5	70	38.19	7.84
6	71	26.83	3.24
7	73	10.11	0.04
8	75	1.39	4.84
9	76	0.03	10.24
10	78	3.31	27.04
11	101	24.82	795.24
12	110	1143.79	...
n	11	11	10
Sum	838	2002.40	1335.60
Mean $\overline{X}$	76.18		
Mean $\overline{X}_n$	72.80 (excluding observations 49 and 110)		

Note: $G_n = 1335.60/2001.40 = 0.6670$. For $n = 11$ and $\alpha = 0.10$, the critical value (Table F) is 0.5204, hence the largest observation is not an outlier.

3 (Sincich 1986) or 4 (Younger 1979). However, this procedure should not be used for small data sets ($n < 19$) because outliers defined as observations exceeding 4 standard deviations from the mean cannot exist for a small sample size (Shiffler 1988).

DESIGNS FOR COMPARING TWO POPULATIONS

In a given set of objectives, there are several alternative experimental designs as well as methods of statistical analysis. If the objective of the experiment is, for example, to find out whether the addition or substitution of a flavoring compound to the current formulation of a product would change the perceived liking, then we may compare the current formulation against the current formulation with the flavoring added to it. In sensory evaluation, the design to determine the effect of flavoring on liking is the well-known paired comparison design. It is the most widely used design in sensory and consumer testing work. The other design to be discussed in this chapter is the group comparison design.

2.1 PAIRED COMPARISON DESIGN

The layout of the paired design is shown in Table 2.1-1. Returning to our example, let X_{11}, X_{12}, .., X_{1n} be the observations from the current formulation and X_{21}, X_{22}, .., X_{2n} the observations from the same formulation but with the added flavoring. These observations are random variables; i.e., they can assume any value of the rating scale used in the evaluation of each formulation. Although the experimental materials are not necessarily paired, the observations (X_{1i}, X_{2i}) are paired in the sense that they come from the same panelists, thus they are correlated and have the property of a paired observation. In the flavoring example, let μ_1 denote the population mean for the current formulation and μ_2 the formulation with the added flavoring. The null hypothesis for testing the effect of flavoring on liking is

$$H_o: \mu_1 - \mu_2 = 0$$

and the alternative hypotheses may be one of the following:

$$H_a: \mu_1 - \mu_2 \neq 0$$

$$H_a: \mu_1 - \mu_2 > 0$$

$$H_a: \mu_1 - \mu_2 < 0$$

Table 2.1-1
Layout of data for a paired comparison design.

Panelist	X_{1i}	X_{2i}
1	X_{11}	X_{21}
2	X_{12}	X_{22}
...	...	...
...	...	...
n	X_{1n}	X_{2n}

Note that the first H_a is a two-sided hypothesis, whereas the last two are one-sided hypotheses. To determine whether the alternative hypothesis is one-sided or two-sided depends, among other things, on the purpose of the study. Assuming that the observed differences, $d_i = X_{1i} - X_{2i}$, are random samples from the population of differences that is normally distributed with mean μ_d and variance σ_d^2, the Student's paired t statistic is used to test H_0. This is given by

$$\begin{aligned} t &= (\bar{d} - \mu_d)/(\sigma_d/\sqrt{n}) \\ &= \bar{d}/(s_d/\sqrt{n}) \end{aligned} \tag{2.1-1}$$

where $\bar{d} = \Sigma d_i/n$

$$s_d = \sqrt{\Sigma(d_i - \bar{d})^2/n - 1}$$

Note that in Eq. (2.1-1), $\mu_d = \mu_1 - \mu_2 = 0$, and the degree of freedom is equal to $n - 1$. For a two-sided test, the null hypothesis is rejected if $|t| > t_{\alpha/2}$, where α is the prescribed significance level of the test. See Table 2.1-2 for the rejection rule of a one-sided test. Consider an example to illustrate the statistical analysis of a paired comparison design.

Example 2.1-1

The data shown in Table 2.1-3 are sensory data from a paired comparison test. Only the data from 10 panelists are shown for illustration. In this table, we want to test whether $\mu_1 - \mu_2 = 0$ against $\mu_1 - \mu_2 \neq 0$, where μ_1 is estimated by the sample mean $\bar{X}_1$ and μ_2 estimated by $\bar{X}_2$. The first step in the analysis is to compute d_i and the deviation of d_i from the mean difference $\bar{d}$. Then $s_d = \sqrt{10.90/9} = 1.101$, and using Eq. (2.1-1) we obtain

$$t = -0.9/0.348 = -2.586.$$

Referring to Table D for 0.05 significance level and 9 DF, we obtain 2.262, which is exceeded by $|t| = 2.586$. Hence, reject the null hypothesis and conclude that treatment X_2 is significantly more well-liked than treatment X_1 at the 5% level.

Table 2.1-2
Rejection rule for the t statistic under various hypotheses.

H_0	H_a	Reject H_0 if:
$\mu_1 - \mu_2 = 0$	$\mu_1 - \mu_2 \neq 0$	$\lvert t \rvert > t_{\alpha/2,DF}$
	$\mu_1 - \mu_2 > 0$	$t > t_{\alpha,DF}$
	$\mu_1 - \mu_2 < 0$	$t < -t_{\alpha,DF}$

There are nonparametric methods to analyze the data in this example that utilize only the signs (plus or minus) of the quantity d_i. These methods are known as the sign test (Dixon and Mood 1946) and the Wilcoxon signed rank test (Wilcoxon and Wilcox 1964) which are discussed and illustrated in Gacula and Singh (1984). However, for sensory/consumer data where tied observations occur in large numbers the paired analysis is recommended over the nonparametric methods.

2.2 GROUP COMPARISON DESIGN

There is another design for comparing two populations, known as the group comparison design. This design is oftentimes used in sensory analysis. A test situation where group comparison design applies is as follows. Two formulations are to be compared through a reference standard. The questionnaire generally used in this situation is given in Fig. 2.2-1. There is no score given to the reference standard, thus the scores for each formulation denote degree of difference between formulations and the reference standard.

Table 2.1-3
Example for calculating the t statistic for paired comparsion design.

Panelist	X_{1i}	X_{2i}	d_i	$(d_i - \bar{d})^2$
1	7	8	−1	0.01
2	6	5	1	3.61
3	5	6	−1	0.01
4	6	7	−1	0.01
5	5	8	−3	4.41
6	7	7	0	0.81
7	5	5	0	0.81
8	7	8	−1	0.01
9	6	7	−1	0.01
10	6	8	−2	1.21
	$\bar{X}_1 = 6.0$	$\bar{X}_2 = 6.9$	$\bar{d} = -0.9$	Total = 10.90

Let X_{1i} and X_{2j}, i = 1, .., n, j = 1, .., m, denote the degree of difference observations for the first (X_1) and the second (X_2) formulations, respectively. In the group comparison design, X_{1i} and X_{2j} are independent random variables with population means μ_1 and μ_2, and variance σ_1^2 and μ_2^2, respectively; μ_1 is estimated by $\overline{X}_1$, μ_2 by $\overline{X}_2$, σ_1^2 by s_1^2, and σ_2^2 by s_2^2. For the group comparison to be effective, we must have small variation in scores among panelists and must have a fairly homogeneous experimental units. The null and the alternative hypotheses for this design are as follows:

$$H_0: \mu_1 = \mu_2 \text{ vs. } H_a: \mu_1 \neq \mu_2$$

$$\text{or } H_a: \mu_1 < \mu_2$$

$$\text{or } H_a: \mu_1 > \mu_2$$

Name _____ Date _____

Product _____ Set no. _____

Sample X is a reference sample. The coded samples *may* or *may not* be different from Sample X. You are looking for *overall flavor* differences only.

Taste Sample X first, then taste each of the coded samples, from left to right. Compare the coded samples against the X sample when making your judgment of degree of overall flavor difference. Make your judgment on *overall flavor differences only* and not appearance or texture differences.

Difference from reference	Sample codes		
No difference			
Very slight difference			
Slight difference			
Moderate difference			
Large difference			
Extremely large difference			
Describe the overall flavor difference, if any: _____			

Fig. 2.2-1
Score sheet used to obtain the data for Example 2.2-1.

Table 2.2-1
Layout of observations for the group comparison design.

	X_{1i}	X_{2j}
	X_{11}	X_{21}
	X_{12}	X_{22}
	...	...
	...	...
	X_{1n}	X_{2m}
Sample size	n	m
Mean	$\overline{X}_1$	$\overline{X}_2$
Variance	s_1^2	s_2^2

If the study objective requires a two-sided test, then the appropriate alternative hypothesis is H_a: $\mu_1 \neq \mu_2$, and so on.

Table 2.2-1 shows the layout of the data. Note that the number of observations between formulations may not necessarily be equal, in contrast to the paired comparison where the number of observations must always be equal being in pairs. The test statistic for group comparison design is also the Student's t statistic computed by

$$t = \overline{X}_1 - \overline{X}_2/s_p\sqrt{(1/n) + (1/n)} \tag{2.2-1}$$

where

$$s_p = \sqrt{[s_1^2(n-1) + s_2^2(m-1)]/(n+m-2)} \tag{2.2-2}$$

The quantity s_p is a pooled standard deviation from both populations. Thus, in the use of the independent t statistic, we must assume that $s_1^2 = s_2^2$, and the observations from both populations must be normally distributed. Large numbers of panelists, over a 100, are generally used in sensory testing so that we may not be concerned about the normality assumption. To test H_0: $\mu_1 = \mu_2$ against one of the alternatives, the computed t with n + m − 2 DF is compared to the tabled t at the desired significance level α. For a two-sided test, reject H_0 if $|t| >$ tabled t value. For a one-sided test, the rejection rules given in Table 2.1-2 apply.

Example 2.2-1

For simplicity, consider a small set of data in Table 2.2-2 obtained by using the questionnaire in Fig. 2.2-1 to illustrate the statistical analysis of group comparison design. Readers may verify the calculations in this table. The pooled standard deviation is found to be

$$s_p = \sqrt{[(1.656)9 + (0.989)9]/18} = 1.150.$$

Table 2.2-2
Example for the statistical analysis of group comparison design.

	X_{1i}	X_{2j}
	3	0
	3	1
	1	1
	4	0
	0	2
	1	3
	1	1
	3	1
	2	2
	1	0
	$n = 10$	$m = 10$
	$\Sigma X_{1i} = 19$	$\Sigma X_{2j} = 11$
	$\bar{X}_1 = 1.9$	$\bar{X}_2 = 1.1$
	$s_1^2 = 1.656$	$s_2^2 = 0.989$
$s_1^2 = [\Sigma X_{1i}^2 - (\Sigma X_{1i})^2/n]/n-1 = 1.656$		
$s_2^2 = [\Sigma X_{2i}^2 - (\Sigma X_{2j})^2/m]/m-1 = 0.989$		

Using Eq. (2.2-1)

$$t = (1.9 - 1.1)/[1.150\sqrt{0.200}] = 1.556$$

is obtained, which is not significant at the 5% level. Therefore, the null hypothesis is "accepted," and it is concluded that the mean of formulation X_1 is not significantly different from the mean of the other formulation.

It should be added that in this example, the basis of comparison or the frame of reference is the standard sample. In monadic testing where the frame of reference is the panelists themselves, the data are analyzed in the same manner as in this example.

COMPLETELY RANDOM AND RANDOMIZED COMPLETE BLOCK DESIGN

In this chapter, the design and statistical analysis used for comparing more than two populations are discussed. The completely random (CR) design is discussed first.

3.1 COMPLETELY RANDOMIZED DESIGN

The CR design is an extension of the group comparison design discussed in the previous section. Therefore, for the CR design to be effective we must also assume that the experimental materials to be used in the experiment are homogeneous. We must also assume that panelists are uniform in their ratings, which is a difficult assumption to meet. These assumptions limit the use of the CR design in sensory testing. However, in some situations when we are limited in our options and must use the monadic test, then the CR design finds its use. To counteract the variation due to panelists, a large number (larger than 100) of panelists is generally required. The randomization process for the CR design consists of random assignment of the panelists to treatment groups. The statistical model is the same as that of Eq. (1.5-1) which is

$$X_{ij} = \mu + T_i + E_{ij}, \qquad \begin{matrix} i=1, .., k \\ j=1, .., n \end{matrix} \tag{3.1-1}$$

where X_{ij} the observed value for the ith treatment and jth panelist, T_i the effect of the ith treatment, and E_{ij} are random errors assumed to be normally and independently distributed with variance σ_e^2. The variance σ_e^2 includes the variation due to panelists, experimental materials, and other errors not controlled by the design.

The layout of the CR design is shown in Table 3.1-1. In this table $X_{1.}$ refers to the total of observations for treatment 1, and so on; $X_{..}$ is obviously the grand total. The analysis of variance table is given in Table 3.1-2 along with the formulas for calculating the sums of squares.

Table 3.1-1
Layout for a completely random design.

	Treatment				
	T_1	T_2...		T_i	
	X_{11}	X_{21}	...	X_{i1}	
	X_{12}	X_{22}	...	X_{i2}	
	...	...	...	...	
	...	...	...	...	
	X_{1j}	X_{2j}	...	X_{ij}	
Total	$X_{1.}$	$X_{2.}$	...	$X_{i.}$	$G = X_{..}$

Example 3.1-1

Table 3.1-3 shows smoke intensity data from a monadic test for three formulations of a food product. The sensory analyst wanted to find out whether there are significant differences among formulations with respect to the level of smoke using a 7-point scale where 1 denotes "zero amount" and 7 denotes "extremely strong" amount. In this table, the calculations of sums of squares are also shown following the formulas outlined in Table 3.1-2. The ANOVA table is given in Table 3.1-4. Recall from Chapter 1 that MS was obtained by SS/DF. The F ratio 6.70/1.56 = 4.29 with 2 numerator DF and 24 denominator DF is significant at the 5% level (Table A).

To determine which means are significantly different from each other, the Duncan's multiple range test at the 5% level is conducted. First, the standard error of the mean S_x is computed:

$$S_x = \sqrt{MSE/n} = \sqrt{1.56/9} = 0.416.$$

To facilitate the calculation of R_p, tabulate the following with the aid of Appendix Table B at $\alpha = 0.05$ and DF = 24:

p :	2	3
$r_{0.05,p,24}$:	2.92	3.07
R_p :	1.21	1.28

Table 3.1-2
ANOVA for the completely randomized design assuming equal number of replications.

Source of Variance	DF	Sums of squares
Total	N − 1	$SSTO = \Sigma X_{ij}^2 - G^2/N, \quad N = nk$
Between treatments	k − 1	$SST = [(X_{1.}^2 + X_{2.}^2 + \ldots + X_{j.}^2)/n] - G^2/N$
Error	N − k	SSE = SSTO − SST

Table 3.1-3
Smoke intensity data for three bacon formulations obtained using a 7-point scale.

	Formulations			
	1	2	3	
	5	7	4	
	4	6	5	
	6	4	7	
	5	7	4	
	5	5	4	
	7	4	3	
	4	7	3	
	3	7	4	
	4	7	5	
Total	43	54	39	G = 136
Mean	4.8	6.0	4.3	

Calculations:
SSTO = $[(5)^2 + (4)^2 + \ldots + (4)^2 + (5)^2] - (136)^2/27 = 50.96$
SST = $[(43)^2 + (54)^2 + (39)^2]/9 - (136)^2/27 = 13.41$
SSE = $50.96 - 13.41 = 37.55$

Using Eq. (1.6-1) in Chapter 1, the R_p is obtained. The pairwise differences between means are as follows noting that $X_2 > X_1 > X_3$:

$$d_{23} = 6.0 - 4.3 = 1.7,\ p = 3$$
$$d_{21} = 6.0 - 4.8 = 1.2,\ p = 2$$
$$d_{13} = 4.8 - 4.3 = 0.5,\ p = 2$$

Since $d_{23} > R_p = 1.28$, this difference is declared significant. The remaining differences did not exceed the R_p of 1.21, hence they are declared to be not significant.

Table 3.1-4
ANOVA table for Example 3.1

Source of Variance	DF	SS	MS	F-ratio
Total	26	50.96		
Between formulations	2	13.41	6.70	4.29
Error	24	37.55	1.56	

3.2 RANDOMIZED COMPLETE BLOCK DESIGN

We now discuss the randomized complete block design (RCB). This is the most widely used design when there are more than two variables or treatments to be compared. The RCB is similar to the paired comparison design in the sense that the experimental materials within rows are assumed to be homogenous. The fulfillment of this assumption will result in improved estimate of treatment effects and random errors, because all the treatments within rows are compared in the same experimental conditions. In the RCB design, the rows are known as blocks and columns as treatments. In statistical terminology, the control of extraneous variation due to rows is known as blocking, and in the case of the paired comparison design it is known as pairing. The use of blocking and pairing for the reduction of extraneous variation is seen often in sensory evaluation work. Panelists generally do not use the scale uniformly, some panelists are high raters and some are low raters. By using panelist as block, all comparisons are within panelist, which would be more precise because panelist variation does not enter in the calculation of treatment differences.

The statistical model in (3.1-1) can be rewritten to include block effects as follows:

$$X_{ij} = \mu + T_i + B_j + E_{ij}, \qquad \begin{matrix} i=1,2, .., k \\ j=1,2, .., n \end{matrix} \qquad (3.2\text{-}1)$$

where X_{ij}, μ, and T_i are as defined by (3.1-1); B_j is the effect of blocks, $\Sigma B_j = 0$, and E_{ij} are random errors assumed to be independently and normally distributed; these random errors would include variations not accounted for by the model such as the interaction between treatments and blocks. The interaction effect can be included in the model if panelists replicate evaluations are available. Table 3.2-1 shows the layout for a RCB design without replication, and the corresponding ANOVA table given in Table 3.2-2. The order of treatment evaluation within blocks (panelists) should be randomized to minimize carry-over effects.

Table 3.2-1
Layout for a randomized complete block design.

Block	Treatment				Total
	1	2	...	k	
1	X_{11}	X_{21}	...	X_{k1}	$X_{.1}$
2	X_{12}	X_{22}	...	X_{k2}	$X_{.2}$
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...
n	X_{1n}	X_{2n}	...	X_{kn}	$X_{.n}$
Total	$X_{1.}$	$X_{2.}$	...	$X_{k.}$	G

Table 3.2-2
ANOVA table for randomized complete block design assuming equal number of observations per treatment.

Source of Variance	DF	Sums of Squares
Total	$N-1$	$SSTO = \Sigma X_{ij}^2 - CF, \quad CF = G^2/N$
Treatments	$k-1$	$SST = [(X_{1.}^2 + X_{2.}^2 + .. + X_{k.}^2)/n] - CF$
Blocks	$n-1$	$SSB = [(X_{.1}^2 + X_{.2}^2 + .. + X_{.n}^2)/k] - CF$
Error	$N-k-n+1$	$SSE = SSTO - SST - SSB$

Example 3.2-1

The data in Example 3.1-1 will be used to illustrate the statistical analysis of RCB design. In this design each row represents a panelist. From Table 3.1-3, it was found that SSTO = 50.96 and SST = 13.41. To compute SSB, row totals are needed and are found to be:

$$X_{.1} = 16 \qquad X_{.2} = 15 \qquad X_{.3} = 17$$
$$X_{.4} = 16 \qquad X_{.5} = 14 \qquad X_{.6} = 14$$
$$X_{.7} = 14 \qquad X_{.8} = 14 \qquad X_{.9} = 16$$

Also, $CF = (136)^2/27 = 685.04$. Then using the formulas in Table 3.2-2,

$$SSB = [(16)^2 + (15)^2 + .. + (16)^2]/3 - CF = 3.63,$$

and $$SSE = 50.96 - 13.41 - 3.63 = 33.92.$$

The ANOVA table is shown in Table 3.2-3. The F ratio 6.71/2.12 = 3.17 did not exceed the tabled F = 3.63 with 2 numerator DF and 16 denominator DF. Hence, the null hypothesis that there are no significant differences among treatments is "accepted." This finding contradicts the result by the ANOVA using the CR design analysis. This is probably due to the reduction in the error DF with a corresponding small reduction in SS error. Note that had the block SS comprised a large portion of the error SS, we may have obtained the same result. Furthermore, the DF is too small to obtain a valid comparison between the designs. The standard error of the mean is $S_{\bar{x}} = \sqrt{2.12/9} = 0.49$.

Example 3.2-2

Suppose that a ranking method was used in Example 3.1-1 to evaluate the three product formulations. A rank of 1 was given to the sample with the "least smoke

Table 3.2-3
ANOVA table for Example 3.2-1

Source of Variance	DF	SS	MS	F ratio
Total	26	50.96		
Treatments	2	13.41	6.71	3.17
Blocks	8	3.63	0.45	0.21
Error	16	33.92	2.12	

intensity'' and a rank of 3 for the sample with the ''most smoke intensity.'' The rank data and the sum of ranks for each formulation are shown in Table 3.2-4. The pairwise differences d_{ij} are shown at the bottom of this table. Note that in d_{ij}, i refers to the sample with the larger rank sum and j for the sample with the smaller rank sum in a given pair comparison. For $k = 3$ and $n = 9$, the critical value for the rank sum multiple comparison test at the 5% level is equal to 10 (Table C). As shown by the d_{ij}s at the bottom of Table 3.2-4, no pairwise comparisons exceeded the critical value of the test. Hence, none of the comparisons are significant at the 5% level, although the comparison between formulations 2 and 3 is significant at the 10% level.

Table 3.2-4
Smoke intensity data for three formulations based on rank scale.

Panelist	Formulation		
	1	2	3
1	2	3	1
2	1	3	2
3	2	1	3
4	2	3	1
5	2	3*	1
6	3	2	1
7	2	3	1
8	1	3	2
9	1	3	2
Rank sum	16	24	14
d_{ij}:	$d_{21} = 24 - 16 = 8$		
	$d_{23} = 24 - 14 = 10$		
	$d_{13} = 16 - 14 = 2$		

* Rank based on mean score due to tied observations.

INCOMPLETE BLOCK DESIGNS

In Chapter 3, the complete block design, in which the treatments appeared in each block, was discussed. In this chapter we discuss a design in which not all treatments will appear in each block.

When the number of treatments to be compared in a sitting is large, the panelists would no longer be effective in their evaluations. Therefore, it is desirable that the number of treatments to be evaluated at one time should be reduced. This can be accomplished by using incomplete block designs. The applications of incomplete block designs to sensory analysis include the papers by Hanson *et al.* (1951), Gacula and Kubala (1972), Gacula (1978), and Chambers *et al.* (1981). In this chapter, the balanced incomplete block (BIB) design and the BIB design augmented with control or reference sample in every block are considered. The readers may refer to Gacula and Singh (1984) for a comprehensive discussion of the applications of incomplete block designs to sensory analysis.

4.1 BALANCED INCOMPLETE BLOCK DESIGN

The rationale behind the BIB design is an extension of the RCB design, that is, to compare treatments within blocks in order to minimize extraneous variation in the comparison. According to Yates (1936), a BIB design has the following parameters:

t = number of treatments,
k = number of experimental units per block,
r = number of replications of each treatment,
b = number of blocks (panelists),
λ = number of blocks in which each pair of treatments are compared.

These parameters are not independent; hence, the use of BIB design must satisfy two equalities, namely, $rt = bk = N$ and $(t - 1) = r(k - 1)$, where N is the total number of observations in the experiment. A basic BIB design satisfies the above equalities which can be verified in the layout in Table 4.1-1 for $t = 4$, $k = 2$, $b = 6$, and $\lambda = 1$. Note that $r = (t - 1)/(k - 1) = 3$; $N = 3(4) = 6(2) = 12$.

Notations for block totals $X_{.j}$, treatment totals $X_{i.}$, and others are also given in Table 4.1-1 to aid in the computation of sums of squares for ANOVA.

The basic BIB design is generally repeated p times to increase the number of replications. Therefore, r, b, and λ are multiplied by p. Cochran and Cox (1957) gave an extensive tabulation of BIB designs, and some designs that are useful in sensory evaluation are given in Appendix Table M. A theoretical discussion of BIB design is given by John (1971, 1980).

The statistical model for BIB design is

$$X_{ij} = \mu + T_i + B_j + E_{ij}, \qquad \begin{matrix} i = 1, 2, .., t \\ j = 1, 2, .., b \end{matrix} \tag{4.1-1}$$

where X_{ij}, μ, T_i, B_j and E_{ij} are similarly defined as in Eq. (3.2-1). However, due to incomplete blocking the sum of squares for treatments is computed differently from that of the RCB design. The procedure for estimating treatment effects, adjusted for block effects, is known as the intrablock analysis. Some information on treatment effects can be also found from the variation between blocks, and the procedure for recovering the information on treatment effects is known as the interblock analysis (Yates 1940). If we can assume that block effects are random effects, the estimates of treatment effects can be improved by combining the intra- and interblock analyses. This assumption is supported by Lundahl and McDaniel (1988). The procedure for the combined analysis is discussed by Gacula and Kubala (1972), Cornell (1974), Seshadri (1963), and Gacula and Singh (1984). Only the intrablock analysis is discussed here. Table 4.1-2 displays the intrablock ANOVA table for the model given by Eq. (4.1-1), along with formulas for computing various sums of squares. In this table, the quantity kQ_i is determined by

$$kQ_i = kX_{i.} - B_{(i)}, \qquad i = 1,2, \ldots, t \tag{4.1-2}$$

where $B_{(i)}$ is the total of all observations in the blocks that contain the ith treatment. Estimates of the treatment means adjusted for block effects are given by

$$X_i = \mu + t_i, \qquad i = 1,2, \ldots, t \tag{4.1-3}$$

where $\mu = G/N$ and $t_i = kQ_i/t\lambda$. The standard error of X_i is

$$SE(X_i) = \sqrt{k(t-1)MSE/t(k-1)r}\,. \tag{4.1-4}$$

Note that $MSE = SSE/(N - b - t + 1)$.

Table 4.1-1
Layout for a balanced incomplete block design.

	Treatments				
Blocks	1	2	3	4	$X_{.j}$
1	X_{11}	X_{21}			$X_{.1}$
2	X_{12}		X_{32}		$X_{.2}$
3	X_{13}			X_{43}	$X_{.3}$
4		X_{24}	X_{34}		$X_{.4}$
5		X_{25}		X_{45}	$X_{.5}$
6			X_{36}	X_{46}	$X_{.6}$
$X_{i.}$	$X_{1.}$	$X_{2.}$	$X_{3.}$	$X_{4.}$	G
$kX_{i.}$	$kX_{1.}$	$kX_{2.}$	$kX_{3.}$	$X_{4.}$	
$B_{(i)}$	$B_{(1)}$	$B_{(2)}$	$B_{(3)}$	$B_{(4)}$	
kQ_i	kQ_1	kQ_2	kQ_3	kQ_4	

Table 4.1-2
Intrablock ANOVA for balanced incomplete block design.

Source of Variance	DF	Sums of Squares
Total	$N - 1$	$SSTO = \Sigma\Sigma X_{ij}^2 - CF$, $CF = G^2/N$
Blocks	$b - 1$	$SSBL = (\Sigma X_{.j}^2/k) - CF$
Treatments adj. for blocks	$t - 1$	$SST = \Sigma kQ_i^2/t k\lambda$
Error	$N - b - t + 1$	$SSE = SSTO - SSBL - SST$

Example 4.1-1

Consider the sensory data given in Table 4.1-3 for $t = 4$, $k = 2$, and $p = 2$. The sums for panelists, treatments, and repetitions can be easily verified. Using treatment 1, let us illustrate the calculations of $kX_{1.}$, B_1, and kQ_1 as follows.

$$kX_{1.} = 2(32) = 64$$

$$B_{(1)} = X_{.1} + X_{.2} + X_{.3} + X_{.7} + X_{.8} + X_{.9}$$
$$= 10 + 13 + 9 + 13 + 13 + 9 = 67$$

$$kQ_1 = 64 - 67 = -3$$

Using the formulas given in Table 4.1-2, the various sums of squares are obtained as shown below.

Table 4.1-3
Sensory data and calculations for Example 4.1-1.

	Treatment					
Panelist	1	2	3	4	$X_{.j}$	Rep. sum
1	5	5			10	
2	7		6		13	
3	5			4	9	
4		6	7		13	
5		6		4	10	
6			8	6	14	69
7	6	7			13	
8	5		8		13	
9	4			5	9	
10		7	7		14	
11		6		5	11	
12			7	4	11	71
$X_{i.}$	32	37	43	28	140 = G	
$kX_{i.}$	64	74	86	56		
$B_{(i)}$	67	71	78	64		
kQ_i	−3	3	8	−8		

$$CF = (140)^2/24 = 816.67$$

$$SSTO = (5^2 + 7^2 + .. + 5^2 + 4^2) - CF = 35.33$$

$$SSBL = [(10)^2 + (13)^2 + .. + (11)^2 + (11)^2]/2 - CF = 19.33$$

$$SST = [(-3)^2 + (3)^2 + (8)^2 + (-8)^2]/4(2)2 = 9.13$$

Since the basic design was repeated twice ($p = 2$), the variation due to repetition must be estimated so that this variation can be separated from the residual/random error. Thus, the sum of squares due to repetition is computed using the formula

$$SSR = [(R_1^2 + R_2^2)/kb] - CF$$

where R_1 and R_2 are, respectively, the total of observations for the first and second repetitions of the basic design. The formula can be readily extended to more than two repetitions. For our example,

$$SSR = [(69)^2 + (71)^2/2(6)] - CF = 0.16.$$

Note that SSBL must be adjusted because different panelists were used between replications. This adjustment is accomplished by subtracting SSR from SSBL. The adjusted SSBL, denoted by SSBL:R, is

$$\text{SSBL:R} = 19.33 - 0.16 = 19.17$$

with $p(b - 1) = 2(6 - 1) = 10$ DF. Finally

$$\text{SSE} = \text{SSTO} - \text{SSBL:R} - \text{SST} - \text{SSR} = 6.87.$$

Table 4.1-4 shows the ANOVA table. Recall that the MS column is obtained by dividing the SS column by the DF column, i.e., 19.17/10 = 1.92. The F ratio, 1.92/0.76 = 2.53, did not exceed the tabled F value with 10 numerator DF and 9 denominator DF. This indicates that the panelists were uniform in their use of the rating scale. Of main interest is the F ratio for treatments. The F value of 4.00 did exceed the tabled F of 3.86, hence, we conclude that at least two treatment effects or means differ significantly from each other.

We illustrate the calculation of the adjusted treatment means for treatment 1. First the grand mean is computed, which is $\mu = 140/24 = 5.83$. Then $t_1 = -3/4(2) = -0.38$. Using Eq. (4.1-3),

$$\bar{X}_1 = 5.83 - 0.38 = 5.45$$
$$\bar{X}_2 = 5.83 + 0.38 = 6.21$$
$$\bar{X}_3 = 5.83 + 1.00 = 6.83$$
$$\bar{X}_4 = 5.83 - 1.00 = 4.83$$

The standard error of the ith adjusted treatment mean is

$$\text{SE}(X_i) = \sqrt{2(3)0.84/4(1)6} = 0.46.$$

Due to repetition, the number of replications per treatment is 2(3) = 6. A multiple comparison test on the adjusted treatment means may follow, noting that 0.46 is the multiplier for $r_{\alpha,p,DF}$ of the Duncan's multiple range test.

4.2 INCOMPLETE BLOCKS AUGMENTED WITH CONTROL

Data obtained from sensory evaluation techniques are generally correlated with quality; i.e., the score a sample receives is highly dependent on the quality of the sample to which it is compared in a given set or block. In order to anchor the scores,

Table 4.1-4
Analysis of variance for Example 4.1

Source of Variance	DF	SS	MS	F ratio
Total	23	35.33		
Repetitions	1	0.16	0.16	0.21
Panelists within repetition	10	19.17	1.92	2.53
Treatment (Adjusted)	3	9.13	3.04	4.00
Error	9	6.87	0.76	

a control or standard sample should be incorporated in a set of treatment comparisons. In the balanced incomplete block design, not all the treatments are in each block, hence the augmentation of each block with a control should make the comparison of treatments, including the control, more sensitive. The idea of augmenting BIB designs traces back to the independent work of Basson (1959) and Pearce (1960) in plant breeding and field experimentation, respectively. Gacula (1978) described this design for food research applications. The recovery of interblock information for the augmented BIB design is given by Williams and Jones (1979).

The construction of the augmented design is achieved by simply adding the control sample R to blocks of the BIB design in a manner shown in Table 4.2-1, where the letter X denotes the specific treatments that are compared within each block. The order of tasting of the three samples from each treatment within each block is determined at random and not as shown in the layout. In some sensory applications, it may be desirable to have the control sample tasted first and the remaining samples tasted in random order. Note that the addition of the control sample modifies the parameters t and k to t + 1 and k + 1, respectively. The BIB designs are widely catalogued, therefore, the construction of the augmented BIB design is greatly facilitated.

Table 4.2-2 contains the intrablock analysis of variance for an augmented BIB design. Because of incomplete blocking, i.e., not all treatments are contained in a given block (k < t), the treatment totals $X_{i.}$ are adjusted for block effects accomplished by calculating

$$Q_i = X_{i.} - [B_{(i)}/(k + 1)], \qquad i = 1,2, .., t + 1 \tag{4.2-1}$$

where $B_{(i)}$, as before for the BIB design, refers to block totals in which treatment i occurs. The calculation of treatment effects, as well as their standard errors, is obtained by the formulas given by Basson (1959) and are given as Eq. (4.2-2) through (4.2-7) below. The estimate of effect due to the control sample (t_R) is

$$t_R = [(k + 1)/bk]Q_R \tag{4.2-2}$$

where Q_R is also obtained by Eq. (4.2-1), with standard error

Table 4.2-1
Construction of augmented BIB design with control.

Basic design for BIB:

	Blocks (Panelists)						
Treatments	1	2	3	4	5	6	Design parameters:
A	X		X		X		t = 4 k = 2
B	X			X		X	r = 3 b = 6
C		X	X			X	λ = 1 N = bk = 12
D		X		X	X		

Basic design augmented with control sample R:

	Blocks (Panelists)						
Treatments	1	2	3	4	5	6	Design parameters:
R	X	X	X	X	X	X	t + 1 = 5 k + 1 = 3
A	X		X		X		r = 3 b = 6
B	X			X		X	λ = 1 N = b(k + 1) = 18
C		X	X			X	
D		X		X	X		

Table 4.2-2
Intrablock analysis of variance for a BIB design with control.

Source of Variance	DF	Sums of Squares
Total	N − 1	$SSTO = \Sigma X_{ijm}^2 - CF$, i = 1,2,.., t + 1 j = 1,2,.., b m = 1,2,.., p
Repetitions	p − 1	$SSR = \{\Sigma R_m/[b(k+1)]\} - CF$
Panelists within repetition	p(b − 1)	$SSBL{:}R = [\Sigma X_{.j}^2/(k+1)] - SSR - CF$
Treatments (Adjusted)	t	$SST = \Sigma t_i Q_i$
Error	By difference	SSE = SSTO − SSR − SSBL:R − SST

$$SE(t_R) = \sqrt{(k+1)MSE/bk}\ . \tag{4.2-3}$$

The estimate of effect due to the ith treatment is

$$t_i = (Q_i + \frac{Q_R}{t})/[(rk + \lambda)/(k+1)] \tag{4.2-4}$$

with standard error

$$SE(t_i) = \sqrt{(t-1)^2(k+1)MSE/tr(kt-1)} \tag{4.2-5}$$

where MSE is the error mean square obtained from Table 4.2-2. If contrasts between the ith treatment and the control sample are desired, the standard error of the contrast is

$$SE(t_i - t_R) = \sqrt{(k + 1)(k + t - 2)MSE/r(kt - 1)} \quad (4.2\text{-}6)$$

Similarly, the standard error between the contrast of two treatments is

$$SE(t_i - t_j) = \sqrt{2(k + 1)(t - 1)MSE/r(kt - 1)} \quad (4.2\text{-}7)$$

Note that when the basic design is repeated p times, the value of b, r, and λ where these appear in Eq. (4.2-3) through (4.2-7) should be multiplied by p.

The adjusted treatment mean is obtained by $\overline{X}_i = \mu + t_i$, $i = 1, 2, .., t$, where μ is the grand mean excluding the control sample, and the control sample by $\overline{X}_R = \mu + t_R$. Standard multiple comparison procedures, such as the Duncan's multiple range test, may be used to conduct multiple comparisons of effects or the adjusted means.

Example 4.2-1

A storage test comparing three treatments (t = 3) of bacon and a control sample was conducted. Bacon samples were evaluated at various times during the test period using a 7-point rating scale where 1 = no off flavor and 7 = very strong off flavor. Table 4.2-3 shows the data and the calculations of sums of squares.

The initial step in the analysis is to calculate the marginal totals and the quantities $B_{(i)}$ and Q_i, $i = 1,2, .., t + 1$. For example, for treatment 1:

$$B_1 = 7 + 8 + 7 + 8 = 30$$

$$Q_1 = 11 - (30/3) = 1$$

From Eq. (4.2-4),

$$t_1 = [1 + (-4.3333/3)]/(10/3) = -0.1333.$$

To obtain the adjusted treatment means, first calculate

$$\mu = (46 - 11)/12 = 2.92.$$

Then

$$\overline{X}_1 = 2.92 - 0.13 = 2.79,$$

$$\overline{X}_2 = 2.92 - 0.73 = 2.19,$$

Table 4.2-3
Data and calculations for an augmented BIB design with parameters $t + 1 = 4$, $k + 1 = 3$, $b = 3$, $p = 2$, $pr = 4$, $pb = 6$, and $p\lambda = 2$.

Rep.	Panelist	1	2	3	R	$X_{.j}$	R_m
I	1	3	2		2	7	
	2	3		4	1	8	
	3		2	3	2	7	22
II	4	2	2		3	7	
	5	3		4	1	8	
	6		3	4	2	9	24
X_i		11	9	15	11		$G = 46$
B_i		30	30	32	46		
$B_i/k + 1$		10	10	10.6667	15.3333		
Q_i		1	−1	4.3333	−4.3333		$\Sigma Q_i = 0$
t_i		−0.1333	−0.7333	0.8667	−1.0833		

Calculations:
$CF = (46)^2/18 = 117.5556$
$SSTO = (3^2 + 3^2 + .. + 1^2 + 2^2) - CF = 14.44$
$SSR = [(22)^2 + (24)^2]/9 - CF = 0.22$
$SSBL{:}R = (7^2 + 8^2 + .. + 8^2 + 9^2)/3 - SSR - CF = 0.89$
$SST = (1)(-.01333) + (-1)(-0.7333) + (4.3333)(0.8667) + (-4.3333)(-1.0833) = 9.05$
$SSE = 14.44 - 0.22 - 0.89 - 9.05 = 4.28$

$$\overline{X}_3 = 2.92 + 0.87 = 3.79,$$

$$\overline{X}_R = 2.92 - 1.08 = 1.84.$$

The analysis of variance for off flavor is displayed in Table 4.2-4. At 3 numerator DF and 9 denominator DF, the 5% tabled F value is 3.86, which is exceeded by $3.0167/0.4759 = 6.339$. Thus, there is evidence to show that at least two means are statistically different from each other. The standard errors shown in Table 4.2-4 are useful in testing for significance of pairwise comparisons of means and in calculating confidence interval of mean differences. For example, one can use the confidence interval technique to test whether the ith treatment mean is significantly different from the control sample mean. This is accomplished by computing the interval

$$d - Z_{\alpha/2}SE \leq D \leq d + Z_{1-\alpha/2}SE$$

where d = estimated difference between means; $Z_{\alpha/2}$, $Z_{1-\alpha/2}$ = normal deviates from the standard normal distribution (See Table 1.3-2, Chapter 1); SE = standard error appropriately selected from Table 4.2-4; and D = value of the difference between means under the null hypothesis, which is equal to zero. Statistical significance is declared when the interval does not include zero.

Table 4.2-4
Analysis of variance of bacon off-flavor.

Source of Variance	DF	SS	MS	F ratio
Total	17	14.44		
Repetitions	1	0.22	0.22	
Panelists within repetition	4	0.89	0.22	
Treatments (Adjusted)	3	9.05	3.02	6.34
Error	9	4.28	0.48	

Calculations of standard errors:

$SE(t_R) = \sqrt{3(0.48)/12} = 0.35$ (Eq. 4.2-3)

$SE(t_i) = \sqrt{(2)^2(3)0.48/12(5)} = 0.31$ (Eq. 4.2-5)

$SE(t_i - t_R) = \sqrt{3(3)0.48/4(5)} = 0.46$ (Eq. 4.2-6)

$SE(t_i - t_j) = \sqrt{2(3)2(0.48/4(5)} = 0.54$ (Eq. 4.2-7)

For example, consider the mean difference between treatment 1 and the control sample. At the 95% confidence interval,

$$0.95 - 1.96(0.46) \leq D \leq 0.95 + 1.96(0.46)$$

or an interval of (0.05, 1.85). The interval does not include zero, hence the difference between the means for treatment 1 and the control is significant at the 5% level.

CROSSOVER DESIGN

This chapter presents the design and analysis of two-period crossover studies useful in consumer testing work. A crossover design is a plan characterized by the measurement of the response of panelists from the evaluation of two treatments, each treatment being evaluated in sequence. Although there are higher order designs that extend to more than two periods of evaluations (Balaam 1968; Kershner and Federer 1981; Laska *et al.* 1983), their use in sensory testing is not recommended because these designs result in an increased number of samples for evaluation. Furthermore, higher order designs require a longer time interval between sensory evaluations, hence they become impractical under home-use conditions.

Crossover designs are extensively used in clinical trials and in other medical research areas (Brown 1980; Grizzle 1965; Koch 1972; Jones and Kenward 1989).

5.1 CROSSOVER DESIGN IN HOME-USE CONSUMER TESTS

In a typical product home-use study, panelists use one product for certain length of time followed by the other product and vice versa. After each use, a questionnaire is completed by each panelist. The layout of the design is shown in Table 5.1-1. In designing a crossover study, two groups of panelists, denoted by I and II, are formed. Panelists are assigned to the two groups at random. In the first period, Group I uses treatment or product A followed by product B. For the panelists in Group II, they will use the products in the reverse order as shown in Table 5.1-1.

Several statistical models for the analysis of crossover design have been given in the literature and we consider a simple model by Hills and Armitage (1979). Each observation in Table 5.1-1 can be described by a simple model as follows:

$$
\begin{aligned}
&\text{Group I} && \text{A: } y_{1i} = X_{1i} + E_{1i} \\
& && \text{B: } y_{2i} = Y_{2i} + E_{2i} \\
&\text{Group II} && \text{B: } y_{1i} = Y_{1i} + E_{1i} \\
& && \text{A: } y_{2i} = X_{2i} + E_{2i} \\
& && i = 1, 2, .., n
\end{aligned}
\qquad (5.1\text{-}1)
$$

Table 5.1-1
Notation and layout of a crossover design.

Group I			Group II		
Period	Treat.	Observation	Period	Treat.	Observation
1	A	$X_{11}, X_{12}, .., X_{1n}$	1	B	$Y_{11}, Y_{12},.., Y_{1n}$
2	B	$Y_{21}, Y_{22}, .., Y_{2n}$	2	A	$X_{21}, X_{22},.., X_{2n}$

Note: i = 1, 2, .., nth panelist.

where y_{1i} and y_{2i} are designation for responses in periods 1 and 2 ; X_{1i} and Y_{2i} are the observed responses for treatments A and B, respectively, for Group I panelist; Y_{1i} and X_{2i} are the observed responses for treatments B and A, respectively, for Group II panelist; E_{1i} and E_{2i} are the observed responses in the absence of treatment effects for periods 1 and 2, respectively, hence they are estimates of random and systematic variation. In this book, the effect of systematic variation will be called carry-over or order effects.

There are two important assumptions for the above model. First, the absence of treatment by period interaction, that is, the difference between treatments A and B is the same regardless of the period sequence in which they were evaluated. The plots in Fig. 5.1-1 and 5.1-2 illustrate the presence of interaction in the data. Interaction effect is present when the difference between A and B is not the same from period 1 to period 2. On the contrary, Fig. 5.1-3 shows a plot when interaction is absent. Second, the absence of order or carry-over effects. To illustrate the presence of order effects in the data, consider an extreme hypothetical example given in Table

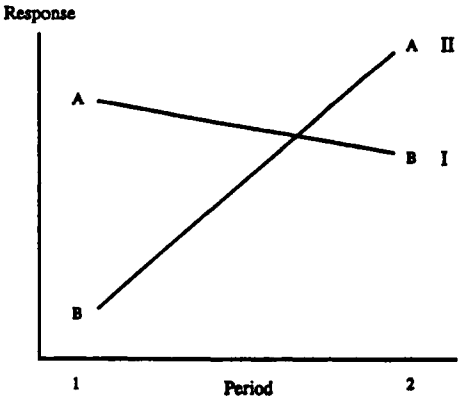


Fig. 5.1-1
Hypothetical outcomes to illustrate the presence of interaction in crossover design. Difference between A and B is larger in period 1 than in period 2.

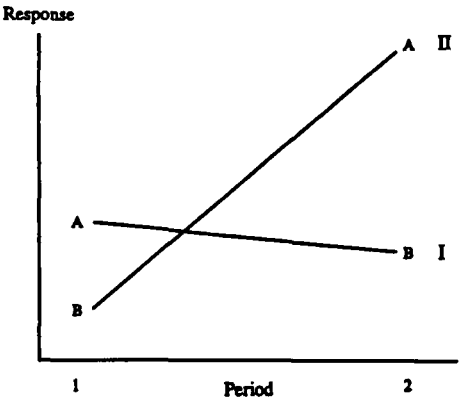


Fig. 5.1-2
Hypothetical outcomes to illustrate the presence of interaction in crossover design. Difference between A and B is larger in period 2 than in period 1.

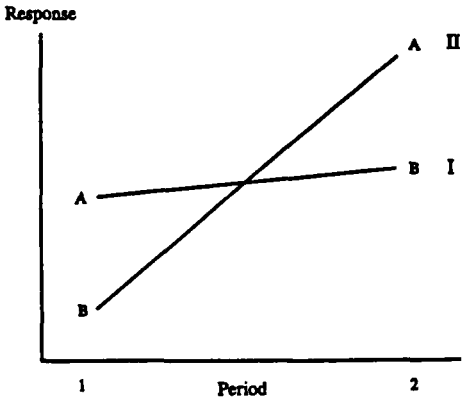


Fig. 5.1-3
Hypothetical outcomes to illustrate the absence of interaction in crossover design. Difference between A and B is the same on both periods.

5.1-2. Product A, with the larger mean score, is favored when it is used in the order (A,B), whereas product B is favored when the order of use was reversed to (B,A). In this example, the mean difference between A and B is zero as a result of order effects. Assuming that treatment effects are nil, this example also illustrates equal

Table 5.1-2
Hypothetical data to illustrate order effects.

	Order of product use		
	(A,B)	(B,A)	Mean
A favored in period 1	A: 6.2	A: 5.0	5.6
B favored in period 2	B: 5.0	B: 6.2	5.6
		Mean difference =	0.0

order effects from A to B and from B to A. It is also possible that treatment and order effects are entangled, and they are said to be confounded or inseparable. Depending on the magnitude of the order effects, their influence on the test of hypothesis generally leads to erroneous acceptance of the null hypothesis. Let us illustrate the estimation and test for significance of treatment and order effects. In this chapter, the estimation of these effects when the response is quantitative, such as a rating scale, and when the response is binary, such as a "yes" or a "no" scale, will be considered.

5.2 RATING SCALE RESPONSE

When ratings such as the hedonic scale are used in the evaluation of treatments, the method of statistical analysis is similar to the paired comparison analysis discussed in Chapter 2. Referring to model Eq. (5.1-1), the estimate of treatment effects is obtained by computing mean differences, $\bar{d}_I$ and $\bar{d}_{II}$, between periods of product use. For Group I panelist

$$d_{Ii} = y_{1i} - y_{2i} = X_{1i} - Y_{2i} + (E_{1i} - E_{2i}) \tag{5.2-1}$$

which estimates the difference between treatments A and B for the ith panelist, i = 1, 2, .., n. Then the mean difference is

$$\bar{d}_I = \Sigma d_{Ii}/n$$

with variance

$$S_I^2 = \Sigma(d_{Ii} - \bar{d}_I)^2/(n - 1) \tag{5.2-2}$$

For the Group II panelist

$$d_{IIi} = y_{2i} - y_{1i} = X_{2i} - Y_{1i} + (E_{1i} - E_{2i}) \tag{5.2-3}$$

which also estimates the difference between A and B. Likewise,

$$\bar{d}_{II} = \Sigma d_{IIi}/n$$

with variance S_{II}^2 calculated using Eq. (5.2-2). The quantity $E_{1i} - E_{2i}$ represents the carry-over effects for the ith panelist. When this quantity is negligibly small, the use of crossover design is unbiased.

It can be shown that the estimate of treatment effects denoted by T_D is

$$T_D = (\bar{d}_I + \bar{d}_{II})/2 \tag{5.2-4}$$

with standard error

$$SE = (1/2\sqrt{(S_I^2/n_I) + (S_{II}^2/n_{II}} \tag{5.2-5}$$

where n_I and n_{II} are the number of panelists in Groups I and II, respectively. Notice that T_D is the average difference of the difference between products or treatments in periods 1 and 2. Consequently, if $\bar{d}_I = \bar{d}_{II}$, then the order effect is zero. Hence, the estimate of order effects denoted by Q is simply

$$Q = (\bar{d}_I - \bar{d}_{II})/2 \tag{5.2-6}$$

with standard error also given by Eq. (5.2-5).

To test for significance of T_D, the two sample t-test is used which is

$$t = T_D/SE \tag{5.2-7}$$

with $n_I + n_{II} - 2$ DF. Replace T_D by Q in Eq. (5.2-7) to test for order effects. In sensory evaluation, it is suggested that the 20% level of significance should be used for testing order effects. Being a preliminary test prior to testing the treatment effects, Grizzle (1965) recommends the 10% level for clinical trials.

Example 5.2-1

Table 5.2-1 shows a data based on the 9-point hedonic scale to illustrate the calculations of treatment and order effects. In this example, Group I panelist used the product in the order (A,B) and Group II panelist used the product in the reversed order (B,A). Note that the differences for order (B,A) in Group II were computed as $y_{2i} - y_{1i}$, i.e., (A − B). As noticed, the calculations in this table follow the paired comparison procedure discussed in Chapter 2, separately applied to Groups I and II. As shown in Table 5.2-1, the estimate of $\bar{d}_I$ is −0.3 and for $\bar{d}_{II}$, 0.4. Using Eq. (5.2-4)

$$T_D = (-0.3 + 0.4)/2 = 0.05$$

Table 5.2-1
Hedonic data and calculations for Example 5.2-1.

Group I (Order A,B)				
Panelist	Period 1 y_{1i}	Period 2 y_{2i}	$d_{Ii} = y_1 - y_2$	$(d_{Ii} - \bar{d}_I)^2$
1	5	6	−1	0.49
2	5	5	0	0.09
3	6	7	−1	0.49
4	5	4	1	1.69
5	6	7	−1	0.49
6	7	8	−1	0.49
7	7	6	1	1.69
8	6	6	0	0.09
9	5	7	−2	2.89
10	7	6	1	1.69
Mean	5.9	6.2	−0.3	Sum 10.10
Group II (Order B,A)				
Panelist	Period 1 y_{1i}	Period 2 y_{2i}	$d_{IIi} = y_{2i} - y_{1i}$	$(d_{IIi} - \bar{d}_{II})^2$
1	6	6	0	0.16
2	7	6	−1	1.96
3	5	5	0	0.16
4	7	8	1	0.36
5	6	8	2	2.56
6	5	7	2	2.56
7	5	5	0	0.16
8	6	5	−1	1.96
9	6	5	1	1.96
10	5	7	2	2.56
Mean	5.8	6.2	0.4	Sum 14.40

Note: A − B = $y_{1i} - y_{2i}$, Group I
A − B = $y_{2i} - y_{1i}$, Group II

as the estimate of treatment effects. The estimate of order effects using Eq. (5.2-6) is

$$Q = (-0.3 - 0.4)/2 = -0.35.$$

Since the estimate of Q is negative, it indicates that the product used second in a comparison is favored. This is clearly shown in Fig. 5.2-1, where product B has higher mean score in period 1 when seen last (order A,B) and lower mean score in period 2 when seen first (order B,A) in the comparison. The same result is observed for product A. Is this result due to chance variation?

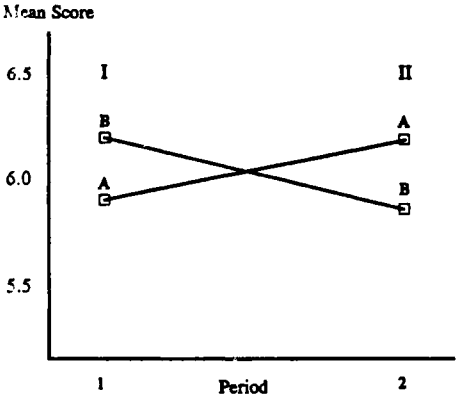


Fig. 5.2-1
Plot of mean scores for Example 5.2-1 to illustrate order effects. Treatments that were seen last have higher mean scores than those seen first.

To test whether T_D and Q are significantly different from zero, the pooled standard deviation for both groups is to be be computed. From Table 5.2-1, the estimate of variance for Group I is 10.10/9 = 1.122 and for Group II, 14.40/9 = 1.600. Using Eq. (2.2-2, Chapter 2), the estimate of the pooled variance is

$$S^2 = [(10 - 1)(1.122) + (10 - 1)(1.600)]/18 = 1.361.$$

Using Eq. (5.2-5)

$$SE = (1/2)\sqrt{(1.361/10) + (1.361/10)} = 0.261.$$

Thus, the t statistic for testing treatment effects is

$$t = 0.05/0.261 = 0.192$$

and for testing order effects

$$t = -0.35/0.261 = -1.341.$$

Each t statistic has 10 + 10 − 2 = 18 DF. The tabled t (Table D) at 18 DF and 5% significance level is 2.101, which is not exceeded by either one of the computed t values (sign of computed t should be ignored). Therefore, the null hypothesis is "accepted" that both treatment and order effects are equal to zero or the effects observed are likely due to chance variation. In other words, the data did not provide

evidence to show that the mean scores for treatments A and B are different, and there is also no evidence that order of product use in this experiment favors either the first or the second product evaluated. If order effect was significant, then there is a problem because T_D is now confounded with Q, and this problem is dealt with in Section 5.4.

5.3 BINARY RESPONSE

When there are only two possible forms of responses, such as "yes" or "no," "preferred" or "not preferred," the resulting outcome is called binary data. For ease of tabulation of results, the responses are coded by the number 1 for a "yes response" and by the number 2 for a "no response." Binary data obtained from a crossover design are summarized in Table 5.3-1. The grand total number of responses is denoted by N. Note also that $N = n_1 + n_2 = n_3 + n_4$. Estimates of treatment effects T_D and order effects Q are obtained following the formula by Nam (1971):

$$T_D = (1/4)\ln(Y_1Y_2/Y_3Y_4) \tag{5.3-1}$$

$$Q = (1/4)\ln(Y_1Y_4/Y_2Y_3) \tag{5.3-2}$$

In Eq. (5.3-2), if the resulting value of Q is positive, the product seen or used first is favored; if Q is negative, the product used second is obviously favored.

To test the null hypothesis for order effects, H_o: $Q = 0$ against its alternative H_a: $Q \neq 0$, a Chi-square statistic with 1 DF is computed as follows.

$$\chi^2 = (\,\}Y_1 - \frac{n_1n_3}{N}\,| - 0.5)^2/V \tag{5.3-3}$$

where

$$V = (n_1n_2n_3n_4)/(N-1)N^2.$$

The null hypothesis of zero order effects is rejected if the computed χ^2 is greater than the tabled χ^2 value (Appendix Table I).

To test the null hypothesis for treatment effect, that is H_o: $T_D = 0$ against H_a: $T_D \neq 0$, use also the formula given by Eq. (5.3-3); however, the tabulation in Table 5.3-1 should be re-arranged as shown in Table 5.3-2. The changes involved the first column and the switching of places for Y_2 and Y_4. Nam (1971) showed that the Gart test (Gart 1969) given by Eq. (5.3-3) remains unbiased for testing treatment effects regardless of the order effects.

Table 5.3-1
Tabulation of binary responses in a two-period crossover design.

	Order of product use		
	(A,B)	(B,A)	Total
A preferred	Y_1	Y_2	n_3
B preferred	Y_3	Y_4	n_4
Total	n_1	n_2	N

Note: Y_1 = number of panelists preferring A in the order (A,B)
Y_2 = number of panelists preferring A in the order (B,A)
Y_3 = number of panelists preferring B in the order (A,B)
Y_4 = number of panelists preferring B in the order (B,A)

Table 5.3-2
Re-arrangement of Table 5.3-1 to accommodate Gart's test for treatment effect.

	Order of product use		
	(A,B)	(B,A)	Total
(1,0)	Y_1	Y_4	n_3
(0,1)	Y_3	Y_2	n_4
Total	n_1	n_2	N

Note: (1,0) = number of times the product used first was preferred.
(0,1) = number of times the product used second was preferred.

In practice, it is recommended to monitor the extent of order effects in sensory testing work. Monitoring of order effects serves as a checkpoint on panelist performance and sensory techniques. A larger number of panelists should be used when order effect is suspected in the data.

Example 5.3-1

A new product had been developed and the research scientist planned to test this newly created product against its number one competitor. A home-use study was proposed using a total of 180 panelists. Half of the panelists, randomly selected, used the products denoted by codes A and B in the order (A,B) and the remaining half, numbering 90 panelists, used the product in the order (B,A). Each product was used during a 7-day period, and after each use a paired comparison questionnaire was completed. The resulting data for overall preference is shown in Table 5.3-3. Note that when product A was used first, 78 panelists preferred A and 12 preferred B. When product B was used first 54 panelists preferred B over A. Clearly,

Table 5.3-3
Overall preference data for products A and B.

	Order of product use		
	(A,B)	(B,A)	Total
A preferred	A: 78	A: 36	114
B preferred	B: 12	B: 54	66
Total	90	90	180

an order effect is indicated. Is this result due to chance? To answer this, the magnitude of order effects is computed using Eq. (5.3-2).

$$Q = (1/4)\ln[(78 \times 54)/(12 \times 36)] = 0.569.$$

To test whether Q is equal to zero, a χ^2 statistic is computed using Eq. (5.3-3). The result is

$$\chi^2 = (\,|\, 78 - \frac{90(114)}{180} \,|\, - 0.5)^2/10.51 = 39.99$$

which is significant at $p < .001$ level (Appendix Table I). This result suggests that the product used first was significantly preferred over the product used second, an indication of the presence of order effects. Although this effect was significant, it was not enough to mask the overwhelming preference for product A as shown by the last column in Table 5.3-3, 114 for product A and 66 for product B. If the difference between products had been small, then the order effects will have a clear impact.

Using Eq. (5.3-1), the estimate of treatment effects is

$$T_D = (1/4)\ln[78 \times 36)/(12 \times 54)] = 0.367.$$

To test H_o: $T_D = 0$, the data are rearranged in the form of Table 5.3-2 resulting in $n_3 = 78 + 54 = 132$. Then the Chi-square statistic is

$$\chi^2 = (\,|\, 78 - \frac{90(132)}{180} \,|\, - 0.5)^2/10.61 = 12.58$$

which is significant at $p < .001$ level. Thus, we conclude that the difference in preference between products A and B is significant in favor of product A.

5.4 ANALYSIS OF DATA WITH CARRY-OVER EFFECTS

Obviously, the presence of carry-over effects is known only after the study has been conducted. If carry-over effects are greatly affecting the estimate of treatment effects, it is possible to reanalyze the data using only the first week data from each group. In this case, we will lose all the second period data from each group in the statistical analysis. The reanalysis of the data is also known as the "first position analysis." The first week data should not contain carry-over effects from either of the products evaluated.

Referring to Eq. (5.1-1), we will only be concerned with comparing the two models shown below:

$$\text{Group I} \quad \text{A: } y_{1i} = X_{1i} + E_{1i}$$

$$\text{Group II} \quad \text{B: } y_{1i} = Y_{1i} + E_{1i}$$

Note that the difference between Groups I and II is a measure of the difference between treatments A and B, free of carry-over effects. To avoid confusion in notation due to the elimination of period 2 data, let us change y_{1i} in Group I to y_{Ii} and y_{1i} in Group II to y_{IIi}. Then the final model is

$$A - B = y_{Ii} - y_{IIi} = (X_{1i} + E_{1i}) - (Y_{1i} + E_{1i})$$

$$i = 1, 2, .., n \text{ panelists}$$

Clearly, the above model resembles the comparison of independent populations, Group I being one population and Group II the other. The statistical analysis is straightforward using the two-sample t-test for group comparison design as discussed in Chapter 2.

There is one point that should be addressed in analyzing only the period 1 data from each group — the sensitivity for detecting differences. The fact that there is no direct comparison between products A and B, certain amount of sensitivity is lost. The evaluation of product A or B is based on the panelist's perception or frame of reference, instead of relative to product A or B. Due to a large variation in the frame of reference among panelists, a certain amount of sensitivity of the test will be lost. Note also that the elimination of the second period data results in a monadic type of data, thus carries with it some concerns in monadic design when not properly used as discussed by Gacula (1987). In practice, it is recommended that the first position analysis should be conducted routinely as a diagnostic tool for detection of order effects and as a check for gross errors in experimental procedures.

Example 5.4-1

Consider the data in Table 5.2-1. Assuming that order effects are large, only the data in period 1 are to be analyzed. The output using the STATISTIX software is given in Table 5.4-1. The result shows that we fail to reject the null hypothesis indicating that there is no significant difference (p = 0.7915) between products A and B at the 5% level.

Table 5.4-1
STATISTIX output for analysis of data in Example 5.4-1.

VIEW DATA

CASE	A	B
1	5.0000	6.0000
2	5.0000	7.0000
3	6.0000	5.0000
4	5.0000	7.0000
5	6.0000	6.0000
6	7.0000	5.0000
7	7.0000	5.0000
8	6.0000	6.0000
9	5.0000	6.0000
10	7.0000	5.0000

TWO SAMPLE T TESTS FOR A VS B

VARIABLE	MEAN	SAMPLE SIZE	S.D.	S.E.
A	5.900	10	8.756E-01	2.769E-01
B	5.800	10	7.888E-01	2.494E-01

	T	DF	P
EQUAL VARIANCES	0.27	18	0.7915
UNEQUAL VARIANCES	0.27	17.8	0.7915

	F	NUM DF	DEN DF	P
TESTS FOR EQUALITY OF VARIANCES	1.23	9	9	0.3804

CASES INCLUDED 20 MISSING CASES 0

FACTORIAL DESIGNS FOR FACTORS AT TWO LEVELS

Factorial designs are experimental plans used to study the effects of two or more factors on process/product attributes, where each level of each factor is varied simultaneously with the other factors in the experiment. Factorial designs are covered in several articles (Finney 1945, 1946; Davis and Hay 1950; McLean and Anderson 1984; Mullen and Ennis 1985) and books (Cochran and Cox 1957; Statistical Eng. Lab. 1957; Box *et al.* 1978; Steel and Torrie 1960; Gacula and Singh 1984; Montgomery 1990a). A widely used factorial design for screening studies is the 2^k factorial design. This design is the foundation of the response surface method discussed in Chapter 8, hence a basic understanding of factorial design is important.

In this chapter, the 2^k design and its fractionation will be discussed to serve as a background material for Chapter 8.

6.1 THE 2^k FACTORIAL DESIGNS

First let us discuss the 2^2 factorial design, followed by the 2^3 and other higher order factorial designs.

The 2^2 Factorial Design

An experiment that consists of k factors A, B, C, .., each at two levels is called a 2^k factorial experiment. The two levels are designated as low and high. In a 2^k factorial design, there are 2^k combinations of factor levels. Each combination can be viewed as a treatment. For a two factor study, k = 2, the total number of combinations is $2^2 = 4$. Denote each combination or treatment by the following symbols:

Treatment:	1	2	3	4
Symbol:	(1)	a	b	ab

In our presentation, a symbol for a factorial combination includes each letter if the corresponding factor is at the high level in the combination. For example, the symbol ab denotes the combination with factors A and B at high level, whereas, the

symbol (1) denotes the combination with all factors at low level. In this notation, the response for each treatment combination is denoted by the letter Y.

The response Y consists of the effects of factors A, B, interaction AB, and the residual E (error). When there is no replication, the AB interaction cannot be estimated. The statistical model is written as

$$Y_{ijl} = \mu + A_i + B_j + (AB)_{ij} + E_{ijl} \qquad (6.1\text{-}1)$$

$$i = 1, 2 \text{ (low, high)}$$

$$j = 1, 2 \text{ (low, high)}$$

$$l = 1, 2, .., r \text{ replications}$$

where Y_{ijl} is the observed response, μ the overall mean in the experiment, A_i the effect of the ith level (high or low) of factor A, B_j the effect of the jth level (high or low) of factor B, $(AB)_{ij}$ the effect of the interaction between factors A and B, and E_{ijl} the random/residual error. The effects are estimated as a deviation from the overall mean as shown in Table 6.1-1. Therefore, the following constraints must hold:

$$\Sigma A_i = 0, \qquad \Sigma B_j = 0, \qquad \Sigma (AB)_{ij} = 0$$

Because of these constraints, one can estimate the residual from the model above by the formula

$$E_{ijl} = Y_{ijl} - [\mu + A_i + B_j + (AB)_{ij}]$$

$$= Y_{ijl} - \hat{Y}_{ijl} \qquad (6.1\text{-}2)$$

which is the difference between the observed and the predicted values. The predicted values are obtained by substituting the observed values into the model (see Example 6.1-1). Most statistical software provides the residual output. In the analysis of fac-

Table 6.1-1
Estimation of effects for the statistical model (6.1-1).

$A_i = Y_{i..} - Y_{...}$
$B_j = Y_{.j.} - Y_{...}$
$AB_{ij} = Y_{ij.} - Y_{i..} - Y_{.j.} + Y_{...}$
$E_{ijl\,k} = Y_{ijl} - Y_{ij.}$

Note: As introduced in Section 1.7 (Chapter 1), the dot notation denotes totals across the subscript replaced by a dot (.).
$Y_{...}$ = Grand mean.
$Y_{i..}$ = Subclass mean response for factor A_i.
$Y_{.j.}$ = Subclass mean response for factor B_j.

torial experiment, factor effects can be evaluated by plotting the residual E_{ijl} against the effects on a normal probability paper. The plotting method is advantageous for higher order factorial and fractional factorial designs because there are more data points that can be plotted. See Example 6.1-2.

In product formulation work, researchers are often faced with studies that involved several factors. Thus, initially the interest would be on screening of these factors by identifying those that have large effects on the desired attributes (response Y). These types of studies are the ideal application of 2^2 factorial design. However, when the number of factors is large, the number of factor combinations increases rapidly. For example, Table 6.1-2 shows the amount of increase in the number of factor combinations with increasing number of factors. The number of factor combinations can be reduced to a manageable size by the use of fractional factorial design. This design is characterized by using only a fraction of the total number of factor combinations to be used in the experiment. In the statistical literature (Finney 1945, 1946) the technique of reducing the number of factor combinations in factorial experiment is known as confounding. This technique will be illustrated in this chapter.

Estimate of Average Factor Effects

Consider a 2^2 factorial experiment with the spatial configuration shown in Fig. 6.1-1. The main effect of factor A is defined as the average change in the response of factor A when its level is changed from high to low across the levels of factor B. In Fig. 6.1-1, the average main effect for factor A is

$$A = \underbrace{[(ab - b)/2]}_{\text{High B}} + \underbrace{[(a - (1))/2]}_{\text{Low B}}$$

$$= (ab - b + a - (1))/2 \qquad (6.1\text{-}3)$$

Table 6.1-2
Total number of factor (treatment) combinations for 2^k factorial design.

Number of factors, k	2^k	Number of factor interactions		
		2-factor	3-factor	4-factor
2	4	2		
3	8	3	1	
4	16	6	4	1
5*	32	10	10	5
6*	64	15	20	15
7*	128	21	35	35
8*	256	28	56	70

*Only up to 4-factor interaction is shown.

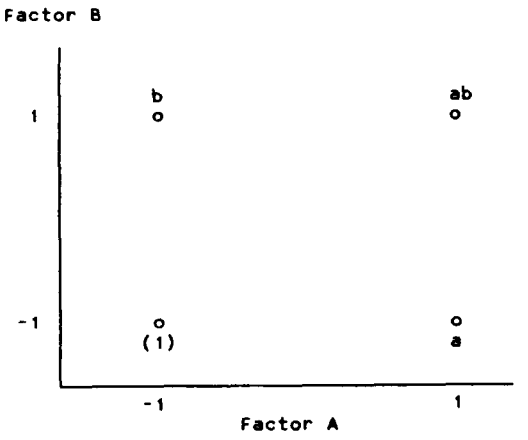


Fig. 6.1-1
Configuration of a 2^2 factorial design.

Note the alternating signs of the treatment combinations, the importance of which will become evident when we discuss design matrix. Also, it can be shown that a half of the average effect is an estimate of regression coefficient for that factor.

Similarly, the estimate of average main effect for factor B is defined as the average change in the response of B when its level is changed from high to low; i.e.,

$$\begin{aligned} B &= \underset{\text{High A}}{[(ab - a)/2]} + \underset{\text{Low A}}{[(b - (1))/2]} \\ &= (ab - a + b - a)/2 \end{aligned} \tag{6.1-4}$$

The estimate of the interaction effect is based on a similar principle, except now it is the average difference in the response of factor A when factor B is at the high and low levels. Thus

$$\begin{aligned} AB &= \underset{\text{High B}}{[(ab - b)/2]} - \underset{\text{Low B}}{[(a - (1))/2]} \\ &= [(ab - b - a + (1)]/2 \end{aligned} \tag{6.1-5}$$

It can also be obtained by computing the average difference in response of factor B when factor A is at the high and low levels as follows.

$$\begin{aligned} AB &= \underset{\text{High A}}{[ab - a)/2]} - \underset{\text{Low A}}{[b - (1))/2]} \\ &= [ab - a - b + (1)]/2 \end{aligned} \tag{6.1-6}$$

If factors A and B do not interact, the difference in response between the high and the low levels of factor B across the levels of factor A will be approximately equal and in the same direction (Fig. 6.1-2). Evidently, the two lines will be parallel. If the two factors interact, the responses will vary with the levels of factor A; hence, the two lines will not be parallel and will often cross, particularly when the interaction effect is large (Fig. 6.1-3).

Note that if each treatment combination is replicated r times the denominator for Eq. (6.1-3) through (6.1-6) is to be multiplied by r. For example

$$A = [ab - b + a - (1)]/2r.$$

Let us summarize the results that we have found for estimating average factor effects in Table 6.1-3. This table contains the signs of the treatment combinations with the coefficient ± 1. This coefficient was omitted in Eq. (6.1-3) through (6.1-6) for simplicity. For example

$$\begin{array}{l} A = [(ab - b) + (a - (1)]/2 \\ \qquad +1 \quad -1 \quad +1 \quad -1 \end{array}$$

Let us introduce a new term called contrast, defined as the difference between the sum of the "+1" responses and the sum of the "−1" responses, i.e.,

$$\text{Contrast} = (\text{Sum of } +1 \text{ responses}) - (\text{Sum of } -1 \text{ responses}).$$

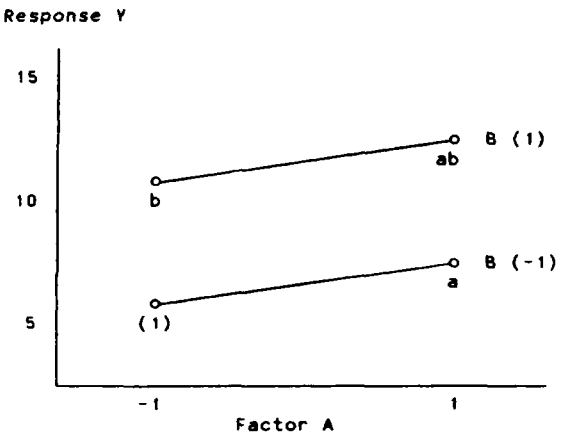


Fig. 6.1-2
Plot to illustrate the absence of interaction between factors A and B.

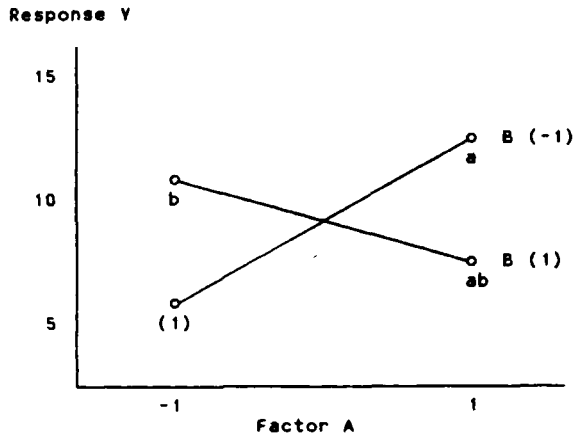


Fig. 6.1-3
Plot to illustrate the presence of interaction between factors A and B.

The coefficients in this table form the design matrix for least squares analysis of the data, using one of the many computer software packages in the market. The coefficients also facilitate analysis by hand calculator as will be illustrated in Example 6.1-1. It also facilitates generalization of the denominator for estimating sums of squares in ANOVA. In the 2^k factorial, the sum of squares is obtained by

$$SS = (\text{contrast})^2/r\Sigma c_i^2 \tag{6.1-7}$$

where c_i is the coefficient of the treatment combinations. In this form, the calculation of SS for fractional factorials becomes easy. It is convenient to present the coefficients of the treatment combinations in tabular form to make the computations of average effects and contrasts easier as shown in Table 6.1-3. The treatment combinations are in the so-called standard order; the rows and columns form the design matrix that plays an important role in computer analysis by the least squares method. For this table, the design matrix is a 4 × 3 (4 rows, 3 columns):

$$\begin{matrix} -1 & -1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ 1 & 1 & 1 \end{matrix}$$

A desirable property of the above matrix is its orthogonality, characterized by the following: the sum of each column is equal to zero and the sum of the products

Table 6.1-3
Useful form for computing average effects and contrasts for a 2 × 2 factorial experiment.

Treatment combinations	Factor effect A	B	AB	ΣY_i
(1)	−1	−1	1	
a	1	−1	−1	
b	−1	1	−1	
ab	1	1	1	
Sum of pluses				
Sum of minuses				
Contrast				
Average effect				

Note: ΣY_i = Sum of observed responses for each treatment combination, i = 1, 2, .., r.
Contrast = (sum of pluses) − (sum of minuses)
Average effect = Contrast/2r
SS = $(\text{Contrast})^2/r\Sigma c_i^2$
SSTO = $\Sigma Y_i^2 - [(\Sigma Y_i)^2/N]$
N = Total number of observations.
SSE = Obtain by difference.

between any two columns is also equal to zero. This property is important to provide a unique solution to least squares equation as illustrated in Chapter 8. In this table, for a complete 2^k factorial design, the divisor of a contrast is $r2^k$ and for the average effect, $r2^{k-1}$.

Example 6.1-1

In this example, two surfactants known to inhibit the growth of bacteria in soaps are evaluated. The first surfactant denoted by factor A is TCC and the second denoted by factor B is Peg-6. The data reported in Table 6.1-4 are percentage of bacterial reduction from a placebo. Note that the responses for each treatment combination were measured at random and not as shown in this table. In fact, the random order was: ab, a, b, (1).

From the information in Table 6.1-4, the average effects of factors A, B, and the AB interaction are:

$$A = 72.7/2(2) = 18.18$$

$$B = 29.3/4 = 7.33$$

$$AB = -10.5/4 = -2.63$$

The sums of squares are:

Table 6.1-4
Data for Example 6.1-1.

Treatment combination	Factor effect A	B	AB	Responses	ΣY_i
(1)	−1	−1	1	13.0, 8.0	21.0
a	1	−1	−1	32.6, 30.0	62.6
b	−1	1	−1	22.4, 18.5	40.9
ab	1	1	1	38.0, 34.0	72.0
Sum pluses	134.6	112.9	93.0		196.5
Sum minuses	61.9	83.6	103.5		
Contrast	72.7	29.3	−10.5		
Ave. effect	18.2	7.3	−2.6		

Note: Contrast for Factor A:
$(62.6 + 72.0) - (21.0 + 40.9) = 72.7$
and so on.
$SSTO = (13.0^2 + 8.0^2 + .. + 38.0^2 + 34.0^2) - [(196.5)^2]/8$
$= 813.24$

$$SSA = (72.6)^2/2(4) = 660.66$$

$$SSB = (29.3)^2/8 = 107.31$$

$$SSAB = (-10.5)^2/8 = 13.78$$

Note that these sums of squares are denoted by TYPE I SS on the SAS output (SAS 1985). The SSE is obtained by difference as

$$SSE = SSTO - (SSA + SSB + SSAB)$$
$$= 813.24 - 781.75 = 31.49$$

If the design matrix is not orthogonal, it would not be possible to subdivide the various SS into component parts as we did in this example. The ANOVA table is shown in Table 6.1-5. The analysis showed that factors A and B significantly affected the amount of bacterial growth on the test material. As shown by the size of the F-ratio, factor A had more effect than factor B; this is also shown by the sizes of their average effects given in Table 6.1-4. The interaction effect is not significant, hence we conclude that the effects of the two factors on bacterial reduction are independent; the average interaction effect was −2.6, which is relatively small compared to the main effects.

Let us use this example to calculate the residuals for model (Eq. 6.1-1). The calculations should provide the reader with a better understanding of what we have discussed in this section. Following the formulas given in Table 6.1-1, the subclass means are as follows:

Table 6.1-5
Analysis of variance for Example 6.1-1.

Source of variance	DF	SS	MS	F-ratio
Total	7	813.24		
A	1	660.66	660.66	83.95**
B	1	107.31	107.31	13.64**
AB	1	13.78	13.78	1.75
Error	4	31.49	7.87	

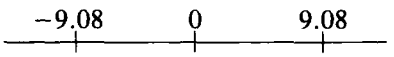
** $p < .01$

	A_1	A_2	Mean
B_1	10.50	31.30	20.90
B_2	20.45	36.00	28.23
Mean	15.48	33.65	24.57

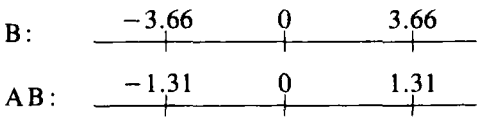
The sum of the effects for factor A is

$$\begin{aligned}\Sigma A_i &= (15.48 - 24.57) + (33.65 - 24.57) \\ &= -9.08 + 9.08 = 0\end{aligned}$$

which satisfies the constraints given earlier. Note that the absolute sum of the above deviations is the average effects of factor A, 9.08 + 9.08 = 18.18, as shown before in this example. Graphically, this looks like this:



It can be shown that half of the average effect is an estimate of the regression coefficient for factor A in the model (Eq. 6.1-1). Likewise, for factor B and interaction AB we have



Therefore, the model is

$$Y = 24.57 + 9.1(A) + 3.7(B) - 1.3(AB)$$

Notice that the sign of the coefficients is determined by the sign of the average effects. Using the GLM procedure of SAS, this model is given by the ESTIMATE

column in the computer output. Using Eq. (6.1-2), the estimate of residual for observation 1 is

$$\begin{aligned}\text{Residual} &= 13.0 - [24.57 + 9.1(-1) + 3.7(-1) - 1.3(1)] \\ &= 13.0 - 10.47 \\ &= 2.53\end{aligned}$$

and so on.

The 2^3 Factorial Design

It is now clear that the 2^3 factorial consists of three factors A, B, and C ($k = 3$) each at two levels, high and low. The number of treatment combinations is $2 \times 2 \times 2 = 8$, and following the same notation as in the 2^2 factorial, the 8 treatment combinations in the standard order are:

$$(1) \quad a \quad b \quad ab \quad c \quad ac \quad bc \quad abc \tag{6.1-8}$$

The configuration of these combinations is given in Fig. 6.1-4. Each combination represents a vertex of a cube. To locate the treatment combinations in this figure, it is useful to first label the vertex for (1), a, b, and c. These vertices would be the origin for finding the other treatment combinations.

The principle for estimating average effects in 2^2 can be easily extended to higher order factorials such as 2^3. There are four comparisons we need to average to obtain the main effect of factor A; these are

$$abc - bc,\ ab - b,\ ac - c,\ a - (1)$$

These comparisons involved the top face and bottom face of the configuration in Fig. 6.1-4. Note that each comparison solves the effect of A by changing its level from high to low. Thus, the average of these four comparisons is

$$A = [abc - bc + ab - b + ac - c + a - (1)]/4 \tag{6.1-9}$$

By arranging the treatment combinations in the standard order, the algebraic sign of each effect is obtained which follow a pattern easily remembered in constructing a design matrix. For example, for factor A

$$\begin{array}{l} A = [-(1) + a - b + ab - c + ac - bc + abc]/4 \\ \qquad\quad -1 \quad +1 \quad -1 \quad +1 \quad -1 \quad +1 \quad -1 \quad +1 \end{array}$$

Similarly, we have the following:

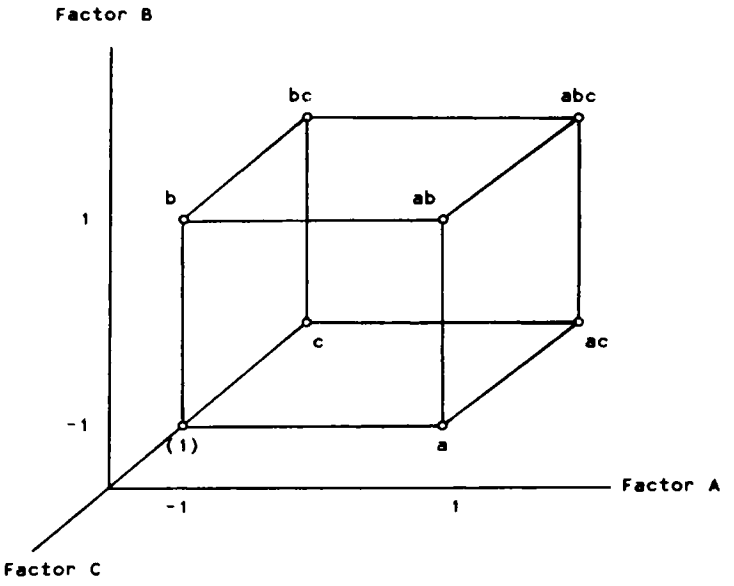


Fig. 6.1-4
Configuration of a 2^3 factorial design.

$$B = [abc - ac + ab - a + bc - c + b - (1)]/4$$

$$C = [abc - ab + bc - b + ac - a + c - (1)]/4$$

Again, refer to Fig. 6.1-4 for a clear picture of these comparisons.

For two-factor interactions, say AB, it involves the front and back faces of the configuration in Fig. 6.1-4. The average AB interaction effect is

$$AB = \underbrace{(abc - bc) - (ac - c)}_{\text{High C}} + \underbrace{(ab - b) - (a - (1))}_{\text{Low C}}/4$$

$$= [abc - bc - ac + c + ab - b - a + (1)]/4$$

Similarly, we have the following:

$$AC = \underbrace{(abc - ab) - (bc - b)}_{\text{High B}} + \underbrace{(ac - a) - (c - (1))}_{\text{Low B}}/4$$

$$= [abc - ab - bc + b + ac - a - c + (1)]/4$$

$$BC = \underbrace{(abc - ab) - (ac - a)}_{\text{High A}} + \underbrace{(bc - b) - (c - (1))}_{\text{Low A}}/4$$

$$= [abc - ab - ac + a + bc - b - c + (1)]/4$$

For the ABC interaction, the average difference between any two-factor interaction for the high and low levels is computed. For example, if we compute the ABC interaction at the high and low levels of factor C, the average ABC interaction is

$$\begin{aligned} ABC &= \underbrace{(abc - bc) - (ac - c)}_{\text{High C}} - \underbrace{(ab - b) + (a - (1))}_{\text{Low C}}/4 \\ &= [abc - bc - ac + c - ab + b + a - (1)]/4 \end{aligned}$$

If we use factor B, we have

$$\begin{aligned} ABC &= \underbrace{(abc - ab) - (bc - b)}_{\text{High B}} - \underbrace{(ac - a) + (c - (1))}_{\text{Low B}}/4 \\ &= [abc - ab - bc + b - ac + a + c - (1)]/4 \end{aligned}$$

which is the same as that using factor C. The same result will be found using factor A.

Example 6.1-2

This example is the same as that of Example 6.1-1, except factor C (Peg-12) is added at 0.5% (low) and 1.0% (high) levels to become a 2^3 factorial. The resulting eight treatment combinations, % bacterial reduction, and calculations are shown in Table 6.1-6. Looking at the average effects, factors A, B, C, and interaction AC appear to be large.

Let us use the normal probability plot to evaluate these effects. The first step is to order the effects from large to small and assign an inverse rank to the effects to obtain the plotting point on the vertical axis. For this example, the plotting point, $(i - 0.5)/7$, is used (See Gacula and Singh 1984). The result is given in Table 6.1-7 and the normal probability plot in Fig. 6.1-5. In the plot, a straight line is drawn on the effects that lie between the 20th and 80th percentiles. Those effects that are far from the line on either direction have large effects on the response. One can immediately see that factors A, B, C, and interaction AC have large effects. Formal test of significance on these effects can be done using the analysis of variance.

An SAS program was written to do the ANOVA and is given in Fig. 6.1-6. In the program the predicted values, residuals, and the plot of the observation versus residual are requested. The resulting SAS output is shown in Table 6.1-8. As shown under the column headed by PR > T, factors A, B, C, and interaction AC were significant. The plot of the residual (Fig. 6.1-7) did not exhibit any particular pattern, suggesting that the observations have homogenous variance, a desirable property to have on a data set.

The column headed by ESTIMATE contains the regression coefficients for each parameter of the statistical model. Thus we have

Table 6.1-6
Data for Example 6.1-2.

	Factor effect									
	A	B	AB	C	AC	BC	ABC	r_1	r_2	ΣY
(1)	−1	−1	1	−1	1	1	−1	13.0	8.0	21.0
a	1	−1	−1	−1	−1	1	1	32.6	30.0	62.6
b	−1	1	−1	−1	1	−1	1	22.4	18.5	40.9
ab	1	1	1	−1	−1	−1	−1	38.0	34.0	72.0
c	−1	−1	1	1	−1	−1	1	16.9	20.1	37.0
ac	1	−1	−1	1	1	−1	−1	10.0	12.0	22.0
bc	−1	1	−1	1	−1	1	−1	22.0	20.0	42.0
abc	1	1	1	1	1	1	1	11.0	15.0	26.0
Sum +	182.6	180.9	156.0	127.0	109.9	151.6	166.5			323.5
Sum −	140.9	142.6	167.5	196.5	213.6	171.9	157.0			
Contrast:	41.7	38.3	−11.5	−69.5	−103.7	−20.3	9.5			

Calculation of sums of squares:

$CF = (323.50)^2/16 = 6{,}540.77$
$SSTO = 7{,}803.39 - 6{,}540.77 = 1{,}262.62$
$SSA = (41.7)^2/2(8) = 108.68$
$SSB = (38.3)^2/16 = 91.68$
$SSC = (69.5)^2/16 = 301.89$
$SSAB = (-11.5)^2/16 = 8.27$
$SSAC = (-103.7)^2/16 = 672.11$
$SSBC = (-20.3)^2/16 = 25.76$
$SSABC = (9.5)^2/16 = 5.64$

Calculation of average effects:

$A = [(1/(4)(2)]41.7 = 5.22$
$B = 38.3/8 = 4.79$
$C = -69.5/8 = -8.69$
$AB = -11.5/8 = -1.44$
$AC = -103.7/8 = -12.96$
$BC = -20.3/8 = -2.54$
$ABC = 9.5/8 = 1.10$

$$Y = 20.22 + 2.60X_1 + 2.39X_2 - 4.34X_3 - 0.72X_1X_2$$
$$- 6.48X_1X_3 - 1.27X_2X_3 + 0.59X_1X_2X_3$$

Notice that the regression coefficients are one-half of the average effects. From this equation we can obtain the residual by substituting the observations into the model to obtain the so-called predicted value of $\hat{Y}$. The residual is simply, observed Y − predicted $\hat{Y}$. Note that we have change notation to conform with regression convention. Here we denote X_1 for factor A, X_2 for factor B, and X_3 for factor C. This notation will be used in Chapter 8.

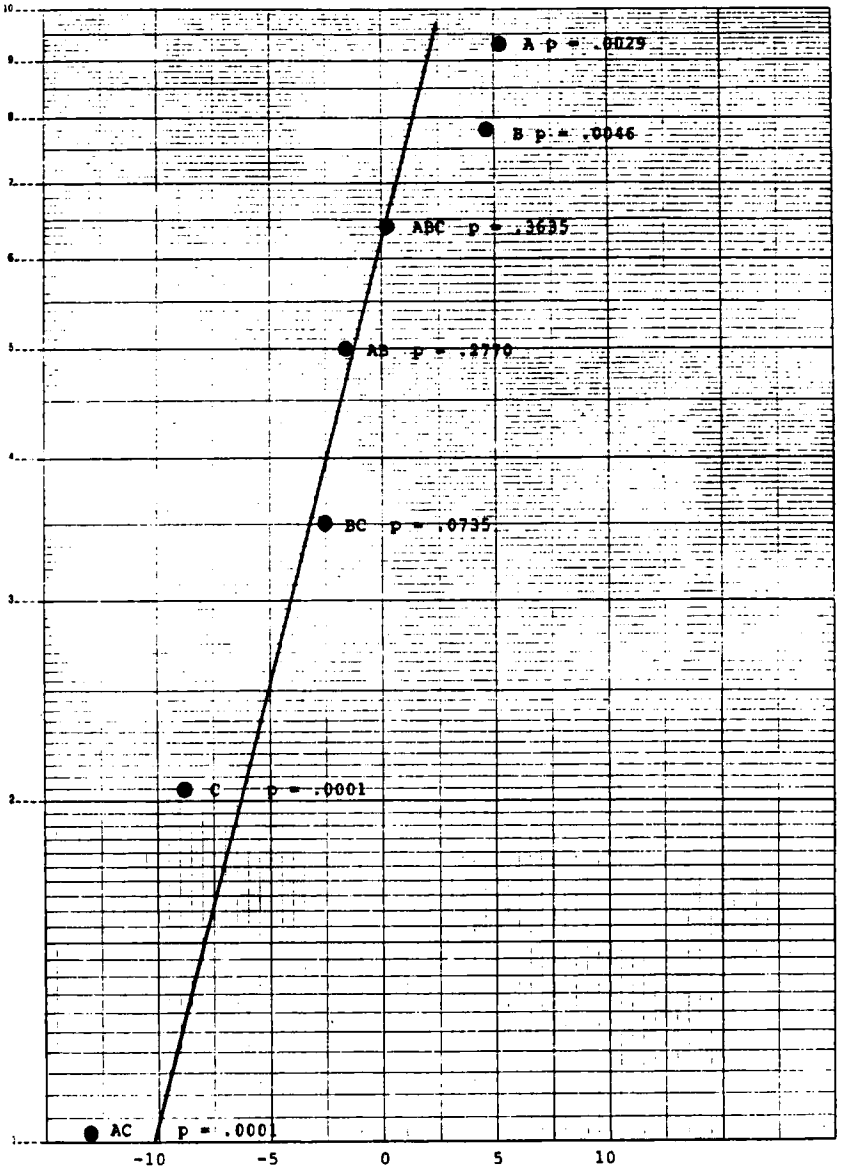


Fig. 6.1-5
Normal probability plot. The p value in this plot is significance probability taken from Table 6.1-8.

Table 6.1-7
Ordered effects and plotting point for normal probability plot (Example 6.1-2).

Factor	Ordered effect	Rank i	(i − .5)/7
A	5.22	7	0.929
B	4.79	6	0.786
ABC	1.19	5	0.643
AB	−1.44	4	0.500
BC	−2.54	3	0.357
C	−8.69	2	0.214
AC	−12.96	1	0.071

Note:
0.929 = (7 − .5)/7
⋮ ⋮
0.071 = (1 − .5)/7

```
**** TSO FOREGROUND HARDCOPY ****
DSNAME=CN0624.AOA.DATA                                    (SAMPLE   )
//AOASAS   JOB  (0624155600GKBI),'GACULA',CLASS=C,TIME=(,15),NOTIFY=AOA,
// MSGCLASS=Z,MSGLEVEL=(0,0)
/*JOBPARM L=010,R=5000,F=9001
/*ROUTE PRINT LOCAL
//STEP1 EXEC SAS,OPTIONS='MACRO NODATE NONUMBER NOSOURCE'
//FT11F001 DD SYSOUT=*,
//            DCB=(BLKSIZE=141,LRECL=137,RECFM=VBA)
//FT12F001 DD SYSOUT=*,
//            DCB=(BLKSIZE=264,LRECL=260,RECFM=VBA)
//*IN   DD    DSN=CN0624.AOA.DATA(PRIMA),DISP=SHR
DATA;
INPUT A B C Y;
CARDS;
 1   -1   -1 13.0
-1   -1   -1  8.0
 1   -1   -1 32.6
 1   -1   -1 30.0
-1    1   -1 22.4
-1    1   -1 18.5
 1    1   -1 38.0
 1    1   -1 34.0
-1   -1    1 16.9
-1   -1    1 20.1
 1   -1    1 10.0
 1   -1    1 12.0
-1    1    1 22.0
-1    1    1 20.0
 1    1    1 11.0
 1    1    1 15.0
PROC GLM;
  MODEL Y = A B C A*B A*C B*C A*B*C;
  TITLE'ANALYSIS OF VARIANCE FOR VARIABLE Y';
  OUTPUT OUT = POUT PREDICTED = PY RESIDUAL = RES;
  PROC PRINT;
  TITLE'PREDICTED AND RESIDUAL ANALYSIS';
PROC PLOT;
  LABEL RES = 'RESIDUAL' Y = '% BACTERIAL REDUCTION';
  PLOT RES*Y = '0';
  TITLE'PLOT OF OBSERVATION AND RESIDUAL';
//
```

Fig. 6.1-6
SAS program to analyze the data of Example 6.1-2. SAS is a registered trademark of the SAS Institute, Inc.

Table 6.1-8
SAS output for a 2^3 factorial experiment: Example 6.1-2.

ANALYSIS OF VARIANCE FOR VARIABLE Y

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	7	1214.01937500	173.43133929	28.55	0.0001	0.961505	12.1911
ERROR	8	48.60500000	6.07562500		ROOT MSE		Y MEAN
CORRECTED TOTAL	15	1262.62487500			2.46487829		20.21875000

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE III SS	F VALUE	PR > F
A	1	108.68062500	17.89	0.0029	1	108.68062500	17.89	0.0029
B	1	91.68062500	15.09	0.0046	1	91.68062500	15.09	0.0046
C	1	801.89062500	49.69	0.0001	1	801.89062500	49.69	0.0001
A*B	1	8.26562500	1.86	0.2770	1	8.26562500	1.86	0.2770
A*C	1	672.10562500	110.62	0.0001	1	672.10562500	110.62	0.0001
B*C	1	25.75562500	4.24	0.0735	1	25.75562500	4.24	0.0735
A*B*C	1	5.64062500	0.93	0.3635	1	5.64062500	0.93	0.3635

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	20.21875000	32.81	0.0001	0.61621957
A	2.60625000	4.23	0.0029	0.61621957
B	2.39375000	3.88	0.0046	0.61621957
C	-4.34375000	-7.05	0.0001	0.61621957
A*B	-0.71875000	-1.17	0.2770	0.61621957
A*C	-6.48125000	-10.52	0.0001	0.61621957
B*C	-1.26875000	-2.06	0.0735	0.61621957
A*B*C	0.59375000	0.96	0.3635	0.61621957

PREDICTED AND RESIDUAL ANALYSIS

OBS	A	B	C	Y	PY	RES
1	-1	-1	-1	13.0	10.50	2.50
2	-1	-1	-1	8.0	10.50	-2.50
3	1	-1	-1	32.6	31.30	1.30
4	1	-1	-1	30.0	31.30	-1.30
5	-1	1	-1	22.4	20.45	1.95
6	-1	1	-1	18.5	20.45	-1.95
7	1	1	-1	38.0	36.00	2.00
8	1	1	-1	34.0	36.00	-2.00
9	-1	-1	1	16.9	18.50	-1.60
10	-1	-1	1	20.1	18.50	1.60
11	1	-1	1	10.0	11.00	-1.00
12	1	-1	1	12.0	11.00	1.00
13	-1	1	1	22.0	21.00	1.00
14	-1	1	1	20.0	21.00	-1.00
15	1	1	1	11.0	13.00	-2.00
16	1	1	1	15.0	13.00	2.00

Y = % bacterial reduction; PY = predicted value of Y
RES = Residual obtained by Y − PY.

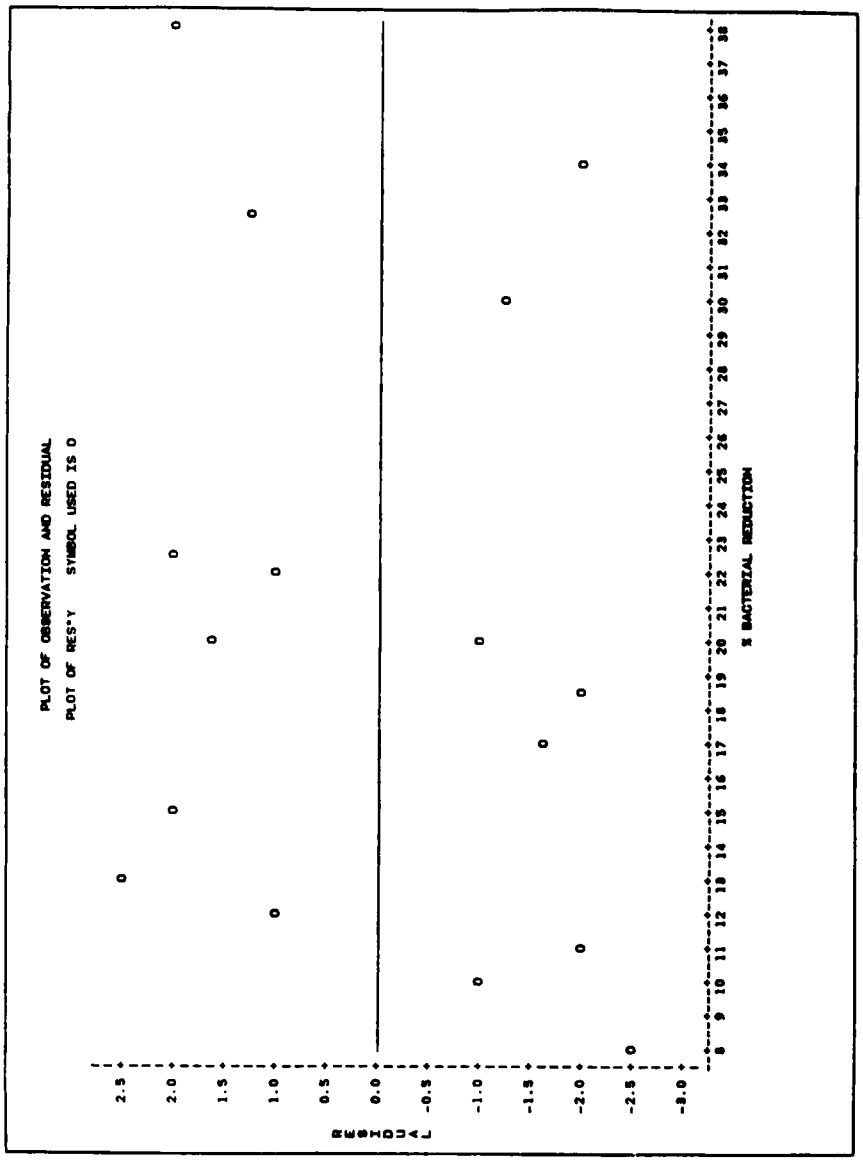


Fig. 6.1-7
Residual plot. Note random distribution of residuals indicating homogeneity of variance in the data.

Addition of Center Point in 2^k Factorial Design

A center point in a design is located halfway between the low and high levels. Fig. 6.1-8 illustrates this for a 2^2 factorial design. The main purpose of the center point is to estimate the lack of fit of the postulated model to the observed data. In Example 6.1-1, we can partition the error sum of squares (SSE) into two parts, one part representing the lack of fit (SSFIT) and the other part representing the so-called pure error (SSPURE). Thus,

$$\text{SSE} = \text{SSPURE} + \text{SSFIT}$$

If the lack of fit is large (statistically significant at an α level), we conclude that one or more effects are not linear, but follow some form of curvature. The SSPURE is obtained from the center point observations, hence we can estimate SSFIT by

$$\text{SSFIT} = \text{SSE} - \text{SSPURE}$$

Note also that SSE can be further improved by replicating the other design points (4 vertices) in Fig. 6.1-8. Table 6.1-9 contains the ANOVA table and relevant formulas for testing the significance of lack of fit. For a numerical example of this topic, see Example 8.1-2 (Chapter 8).

6.2 ONE-HALF FRACTION OF 2^4

Suppose that k = 4 factors denoted by A, B, C, and D. The total number of combinations is $2^4 = 16$. These factor combinations, also called treatments, are as follows:

Treatment	Symbol	Treatment	Symbol
1	(1)	9	d
2	a	10	ad
3	b	11	bd
4	ab	12	cd
5	c	13	abd
6	ac	14	acd
7	bc	15	bcd
8	abc	16	abcd

Recall the interpretation of each symbol as given in the previous section. Suppose that 16 treatments are found to be too large for experimentation, and the researcher has decided on using only a half of them or 8 treatments. How do we go about selecting the 8 treatments to be used? In the 2^4 factorial, there are $2^4 - 1$ effect parameters,

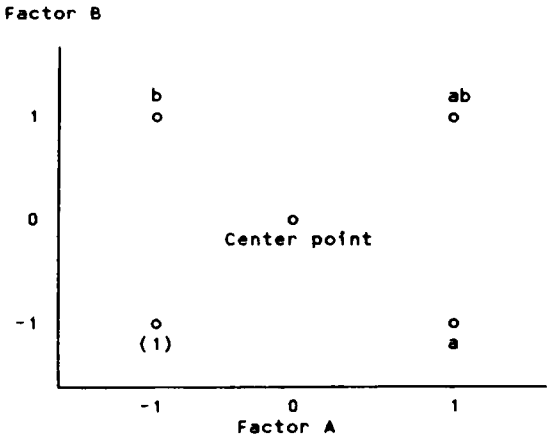


Fig. 6.1-8
Augmentation of a 2^2 factorial design by a center point.

in addition to the grand mean μ, to be estimated. If we design an experiment using only 8 factor combinations, we cannot estimate all the 16 parameters. But we can estimate 8 linear combinations of the effects, each linear combination consisting of only 2 effects. The 2 effects in a combination are completely confounded because their effects cannot be separately estimated. The effect confounded is the price we pay for reducing the number of treatments to be used. However, we have the liberty of choosing the effect to be confounded, and generally we confound the effect that appears to be not important or difficult to interpret. This effect is usually the highest order factor interaction. In this example, the highest order interaction is ABCD.

Suppose we decide to confound the ABCD interaction with μ. If we use the equality sign to denote the confounding and I to denote μ in the equation, then we have a so-called defining contrast:

Table 6.1-9
ANOVA table for 2^2 factorial with center point.

Source of variance	DF	SS	MS	F-ratio
Total	$N - 1$	SSTO		
Regression				
b_1	1			
b_2	1			
Residual	$N - k - 1$	SSE	MSE	
Lack of fit	$M - k$	SSFIT	MSFIT	MSFIT/MSPURE
Pure error	$N - M - 1$	SSPURE	MSPURE	

Note: N = total number of observations including the number of replications at the center point.
k = number of factors, k = 2.
$M = 2^k$.

$$I = ABCD$$

The defining contrast, representing the confounding of μ with the ABCD effect, determines uniquely all the other pairs whose components, called aliases, are confounded. The alias of factor effect A, for example, is obtained from the defining contrast as follows: Multiplying both sides of the defining contrast by A, we find

$$A = A^2BCD, \tag{6.2-1}$$

where in the multiplication operation I is treated as unity. Also, the multiplication is always further simplified by letting $A^2 = I$, $B^2 = I$, and so on. Therefore, Eq. (6.2-1) becomes $A = BCD$, which means that effect A is confounded with BCD interaction. That is, when we measure the effect of A we are actually measuring the combined effects of A and BCD.

Once we have determined the defining contrast, we can write it as a function L,

$$L = l_aX_a + l_bX_b + l_cX_c + l_dX_d \tag{6.2-2}$$

where $l_a = 1$ if factor A appears in the factor combination and 0 otherwise. Likewise, l_b, l_c, and l_d are the indicator functions for factors B, C, and D. The L is reduced to L mod 2. By definition, L mod 2 is the remainder when L is divided by 2. For example, we have the following:

L	L mod 2
1	1
2	0
3	1
4	0
5	1
6	0

Table 6.2-1 shows the results by the application of Eq. (6.2-2). For factor combination (1), all factors at low level, the function L is

$$L = 0(X_a) + 0(X_b) + 0(X_c) + 0(X_d) = 0.$$

For factor combination abcd,

$$L = 1(X_a) + 1(X_b) + 1(X_c) + 1(X_d) = 4,$$

and so on.

Finally, factor combinations corresponding to L mod 2 = 0 and L mod 2 = 1 are separated into two blocks, each block consisting of 8 factor combinations as shown

Table 6.2-1
Construction of a one-half replicate of a 2^4 factorial with ABCD as the defining contrast.

Factor combination	X_a	X_b	X_c	X_d	L	L mod 2
(1)	0	0	0	0	0	0
a	1	0	0	0	1	1
b	0	1	0	0	1	1
ab	1	1	0	0	2	0
c	0	0	1	0	1	1
ac	1	0	1	0	2	0
bc	0	1	1	0	2	0
abc	1	1	1	0	3	1
d	0	0	0	1	1	1
ad	1	0	0	1	2	0
bd	0	1	0	1	2	0
cd	0	0	1	1	2	0
abd	1	1	0	1	3	1
acd	1	0	1	1	3	1
bcd	0	1	1	1	3	1
abcd	1	1	1	1	4	0

in Table 6.2-2. The alias pairs were obtained by the application of the confounding technique discussed earlier in this chapter. It is a matter of choice which of the two blocks to use in the experiment. Note that estimates of main effects (A,B,C,D) are confounded with 3-way interactions, and so on. Suppose we elect to use block 2. To start the experiment, one would formulate 8 treatments defined by block 2. Treatment 1 (or a) would consist of factor A at high level and factors B, C, and D at low level; treatment 2 (or b) would consist of factor B at high level and factors A, C, and D at low level, and so on. The 8 treatments may be too large for simultaneous evaluation, hence we may use the incomplete block design discussed in Chapter 4.

The construction of fractional factorial design for $n > 4$ is given by Kempthorne (1947), Cochran and Cox (1957), Peng (1967), Gacula and Singh (1984), McLean and Anderson (1984), Mullen and Ennis (1985), and recently by Montgomery (1990a).

Example 6.2-1

In this example, the method of statistical analysis of data obtained from a one-half fraction of the 2^4 factorial design is illustrated. As shown previously, there are 8 factor combinations of a half-fraction of 2^4 factorial design. Suppose a BIB design is used to evaluate the 8 treatments ($t = 8$), where each panelist evaluates only 4 treatments ($k = 4$). For $t = 8$ and $k = 4$, 14 panelists are needed to complete a basic BIB design given in Table 6.2-3. One repetition of the basic BIB design constitutes one replication of each factor combination. The BIB adjusted means are the

Table 6.2-2
Aliases and one-half replicate of 2^4 factorial design.

Block 1 (L mod 2 = 0)	Alias pairs	Block 2 (L mod 2 = 1)	Alias pairs
(1)	μ, ABCD	a	A, BCD
ab	AB, CD	b	B, ACD
ac	AC, BD	c	C, ABD
bc	BC, AD	abc	ABC, D
ad	AD, BC	d	D, ABC
bd	BD, AC	abd	ABD, C
cd	CD, AB	acd	ACD, B
abcd	ABCD, μ	bcd	BCD, A

raw data to be used in fractional factorial analysis. Once we have these data for each factor combination, a table is constructed as shown in Table 6.2-4, where two repetitions of the design were done. The row indicated by sum is the total of the adjusted treatment means from the first and second repetitions. The plus coefficient (+1) and the minus coefficient (−1) corresponding to the main effects denotes the presence and the absence of a factor in the factor combination, respectively. For example, factor A has a +1 coefficient for treatments a, abc, abd, and acd; factor B has a −1 for treatments a, c, d, and acd, and so on. The coefficients for the interaction factor effects were obtained by algebraic multiplication of the coefficient of the factors involved in the interaction. For example, for treatment bcd the coeffi-

Table 6.2-3
A basic BIB design for t = 8, k = 4, and b = 14.

Panelist	Treatment number				Corresponding factor combination			
1	1	2	3	4	a	b	c	abc
2	5	6	7	8	d	abd	acd	bcd
3	1	2	7	8	a	b	acd	bcd
4	3	4	5	6	c	abc	d	abd
5	1	3	6	8	a	c	abd	bcd
6	2	4	5	7	b	abc	d	acd
7	1	4	6	7	a	abc	abd	acd
8	2	3	5	8	b	c	d	bcd
9	1	2	5	6	a	b	d	abd
10	3	4	7	8	c	abc	acd	bcd
11	1	3	5	7	a	c	d	acd
12	2	4	6	8	b	abc	abd	bcd
13	1	4	5	8	a	abc	d	bcd
14	2	3	6	7	b	c	abd	acd

Table 6.2-4
Texture data and calculations for Example 6.2-1.

Treatment:	a	b	c	abc	d	abd	acd	bcd			
Repetition 1:	6.9	6.2	6.4	7.2	5.0	7.2	7.8	7.5	Sum of:		
Repetition 2:	7.9	6.9	6.2	7.0	5.5	6.9	7.4	7.0	+	−	Contrast
Sum:	14.8	13.1	12.6	14.2	10.5	14.1	15.2	14.5			
Factor effect											
A(BCD)	+1	−1	−1	+1	−1	+1	+1	−1	58.3	50.7	7.6
B(ACD)	−1	+1	−1	+1	−1	+1	−1	+1	55.9	53.1	2.8
C(ABD)	−1	−1	+1	+1	−1	−1	+1	+1	56.5	52.5	4.0
D(ABC)	−1	−1	−1	−1	+1	+1	+1	+1	54.3	54.7	−0.4
AB(CD)	−1	−1	+1	+1	+1	+1	−1	−1	51.4	57.6	−6.2
AC(BD)	−1	+1	−1	+1	+1	−1	+1	−1	53.0	56.0	−3.0
BC(AD)	+1	−1	−1	+1	+1	−1	−1	+1	54.0	55.0	−1.0

Calculations:

$G = 14.8 + 13.1 + .. + 15.2 + 14.5 = 109$

$CF = (109)^2/16 = 742.56$

$SSTO = [(6.9)^2 + (6.2)^2 + .. + (7.4)^2 + (7.0)^2] - CF$

$= 751.86 - CF = 9.30$

cient for effect BC is the product of the coefficients for B and C, e.g., $(+1)(+1) = +1$. These coefficients are also known as design matrix in the statistical literature.

To determine the sums of squares for each factor effect, contrasts are computed by subtracting the sum of the minuses from the sum of the pluses. For example, for factor A the contrast is

$$(14.8 + 14.2 + 14.1 + 15.2) - (13.1 + 12.6 + 10.5 + 14.5)$$

$$= 58.3 - 50.7 = 7.6$$

and so on. Denoting r as the number of observations per factor combination (treatment), the SS are computed using Eq. (6.1-7). Then

$$SSA = (7.6)^2/2(8) = 3.61$$
$$SSB = (2.8)^2/16 = 0.49$$
$$SSC = (4.0)^2/16 = 1.00$$
$$SSD = (-0.4)^2/16 = 0.01$$
$$SSAB = (-6.2)^2/16 = 2.40$$
$$SSAC = (-3.0)^2/16 = 0.56$$
$$SSBC = (-1.00)^2/16 = 0.06$$

The ANOVA table is summarized in Table 6.2-5. In this table the alias of an effect is shown in parenthesis, i.e., A(BCD). The effects of factors A and C on the hedonic response for texture were found significant. These effects include their respec-

Table 6.2-5
Analysis of variance of a one-half 2^4 fractional factorial with two observations per factor combination.

Source of variance	DF	SS	MS	F ratio
Total	15	9.30		
A (BCD)	1	3.61	3.61	24.07**
B (ACD)	1	0.49	0.49	3.27
C (ABD)	1	1.00	1.00	6.67*
D (ABC)	1	0.01	0.01	0.07
AB (CD)	1	2.40	2.40	16.00**
AC (BD)	1	0.56	0.56	3.73
BC (AD)	1	0.06	0.06	0.40
Error	8	1.17	0.15	

Note:
$Y_{ijlmp} = \mu + A_i + B_j + C_l + D_m + (AB)_{ij} + (AC)_{il} + (BC)_{jl} + E_{ijlmp}$
$i = 1,2;\ j = 1,2;\ l = 1,2;\ m = 1,2;\ p = 1,2$
*$p < .05$; **$p < .01$

tive alias effects, and there is no way of separating them. If our assumption of a negligible 3-way interaction effect holds, then we have confidence on the result of the F test for main effects. The AB interaction effect was also found significant, on the assumption that the CD interaction effect is nil. The assumption of negligible alias effect poses a problem in interpretation of results from fractional factorial experiments. For effective use of fractional factorial design, the nature of the factors being studied should be carefully examined to validate the assumption.

There are other methods for the statistical analysis of fractional factorial experiments. One method is by Yates (1937), which is illustrated in Cochran and Cox (1957), and the other method is the straightforward application of the analysis of variance for multi-classified data. See Kirk (1968) for its application to fractional factorial experiments.

6.3 ONE-HALF AND ONE-FOURTH FRACTION OF 2^k

The procedure for fractionating other 2^k factorial designs, as well as their statistical analysis, follows the procedure given in the preceding section. Two useful fractional factorial plans with $k = 5$ and $k = 6$ are discussed here. In sensory evaluation work the evaluation of a large number of factors or stimuli (greater than 6) is not recommended because of the complex nature of how one perceives various mixtures of stimuli. In addition, an increased number of confounded effects due to fractionation can lead to confusion and problems of interpretation. Other means of decreasing the number of factors, such as a preliminary bench top formulation work and the

standardization of other controllable factors to acceptable levels, may be used. Table 6.3-1 gives the one-half fraction of 2^5 and the one-fourth fraction of 2^6. The factor combinations column in this table is the combinations of factors to be used in the formulation work. Note that in Design 2, the estimate of main effect due to factor A (treatment a) is confounded with 3 aliases, which are the BCE, BDF, and the ACDEF interactions. These aliases were found by the appropriate multiplication

Table 6.3-1
Fractional factorial design for a one-half 2^5 and a one-fourth 2^6.

Design	Factor combinations	Aliases
Design 1:	(1)	ABCDE
k = 5	ae	BCD
1/2 fraction	be	ACD
16 treatments	ab	CDE
Defining contrast is ABCDE	ce	ABD
	ac	BDE
	bc	ADE
	abce	D
	de	BC
	ad	BCE
	bd	ACE
	abde	C
	cd	ABE
	acde	B
	bcde	A
	abcd	E
Design 2:	(1)	ABCE, ABDF, CDEF
k = 6	a	BCE, BDF, ACDEF
1/4 fraction	b	ACE, ADF, BCDEF
16 treatments	c	ABE, ABCDF, DEF
Defining contrasts are ABCE,	d	ABCDE, ABF, CEF
ABDF, and CDEF	e	ABC, ABDEF, CDF
	f	ABCEF, ABD, CDE
	ab	CE, DF, ABCDEF
	ac	BE, BCDF, ADEF
	ad	BCDE, BF, ACEF
	ae	BC, BDEF, ACDF
	af	BCEF, BD, ACDE
	cd	ABDE, ABCF, EF
	cf	ABEF, ABCD, DE
	acf	BEF, BCD, ADE
	acd	BDE, BCF, AEF

of the defining contrasts and the factor effect. For example, with the defining contrasts given by the aliases of (1), factor A has the following aliases:

$$A \times ABCE = A^2BCE = BCE$$

$$A \times ABDF = A^2BDF = BDF$$

$$A \times CDEF = ACDEF$$

and so on. An extensive tabulation of fractional factorial plans is given in several publications (Statistical Eng. Lab. 1957; McLean and Anderson 1984; Mullen and Ennis 1985; Montgomery 1990a).

SCALING METHODS

Measuring subjective responses is a fundamental process involved in sensory evaluation of consumer products, materials and services. The choice of a scale to measure these responses is critical to the outcome and the correct interpretation of the results of experiments. In this chapter, the Thurstone-Mosteller model and ranking procedures for measuring subjective responses are discussed.

7.1 SENSORY MEASUREMENTS

Scaling refers to the processes and techniques used to validate the existence of a defined property of an object or event and to establish operational indices of the relative magnitudes of the property (Gorden 1977). The assignment of numerals to the property of an object is called measurement (Stevens 1946). In this definition, the "defined property" corresponds to product attribute, and "object" refers to product or treatment in the context of consumer testing and experimental design. In this discussion we may refer to products or treatments as stimuli. In the past several decades, scaling methods have proved to be an invaluable tool to sensory analysts for quantifying perception of a given stimulus.

The history of sensory measurements is extremely interesting and well documented over several decades. A brief review of sensory measurements is given by Jones (1974). For our purposes, it is sufficient to discuss briefly the four classes of sensory scales as they are basic to the understanding of sensory measurements. There are four classes of scales: nominal, ordinal, interval, and ratio. The theory and applications of these scales are widely discussed (Stevens 1946, 1951; Guilford 1954; Torgerson 1958; Baird and Noma 1978). It is generally held that the ordinal and the interval scales are the most widely used in consumer testing.

Nominal Scale

The nominal scale is the most simple form that deals with the identification and classification of objects or stimuli. The numbering of football players for identification is a form of nominal scale. Binary responses such as "perceived" or "not perceived," "yes" or "no," used in threshold studies for classifying responses are another example.

Table 7.1-1
Statistical Methods Permissible for the Four Types of Scales.

Scale	Measurement scale value X invariant under mathematical transformation, $Y = f(X)$	Permissible statistics	Example
Nominal	Permutation groups $Y = f(X)$, $f(X)$ means any one-to-one substitution	Number of cases, mode, contingency table correlation	Numbers assigned to players for identification
Ordinal	Isotonic group $Y = f(X)$, $f(X)$ means monotonic function	In addition to foregoing statistics, median, percentiles	Hardness of minerals; grades of leather, wood, etc.
Interval	General linear group, $Y = \alpha + \beta X, \beta > 0$	In addition to all statistics for nominal and ordinal scales, mean, standard deviation, correlation, analysis of variance	Temperature (°F and °C); calendar dates
Ratio	Similarity group, $Y = \beta X, \beta > 0$	In addition to all statistics appropriate for the other three scales, geometric mean, harmonic mean, coefficient of variation	Length, weight, loudness scale

Source: Stevens (1946, 1951, 1961).

Ordinal Scale

The ordinal scale, typified by the ranking method, is popular because of its simplicity. It provides rank order of stimuli, but does not provide meaningful distance of differences because all differences between ranks are obviously equal to 1. Ordinal scale has a true zero point or origin on the scale. In this chapter, we will illustrate rank scaling by paired comparison and by the balanced incomplete block design.

Interval Scale

The interval scale is typified by the hedonic scale developed by Peryam and Girardot (1952). In our discussion, we may use the word category interchangeably with the word interval. Interval scaling is considered a direct method of scaling characterized by a fixed point linear representation of categories on the scale. This scale provides estimates of the distance between stimuli assuming that the difference between scale categories is approximately equal. However, this scale does not possess a true zero point. In practice, the number of categories have ranged from 5 through 10.

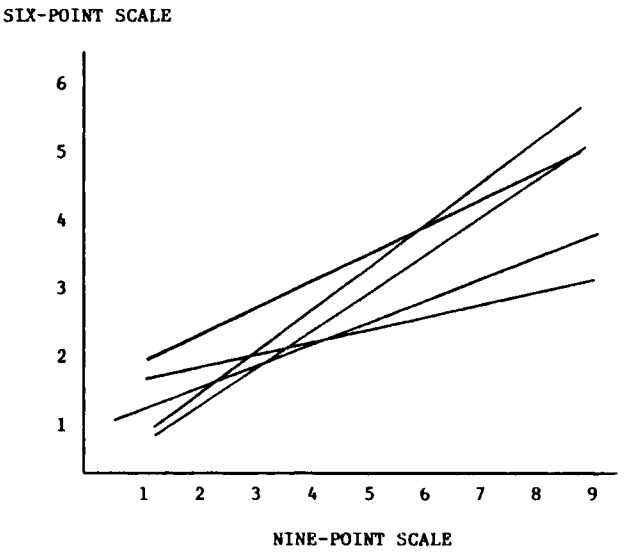


Fig. 7.1-1
Linear relationship between the six-point and the nine-point hedonic scale. Each line is for individual panelist.

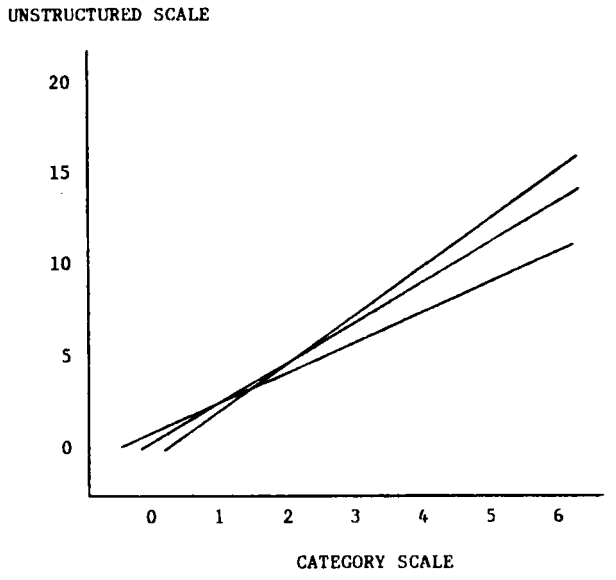


Fig. 7.1-2
Linear relationship between unstructured and category scales to illustrate invariance property. Each line is for individual panelist.

Furthermore, the number of categories has been found to be invariant (Parducci 1982; Gacula 1987), i.e., the scale measures the same property regardless of the length of the scale. Figures 7.1-1 and 7.1-2 illustrate the invariance property of rating scale.

However, one must remember that in general shorter scales (4-point) may not provide enough range for discrimination especially for panelists with high sensory acuity; on the other hand, longer scales (> 11-point) may tend to exaggerate physical differences and destroy scale continuum. Individual skills, experience, and understanding of the sensory attributes being measured will contribute to the successful choice of measurement scales. See Riskey (1986) for a current review of the use of category scales in sensory measurement.

Ratio Scale

The ratio scale approaches a true measurement scale such as a ruler. Ratio scale has a true zero point and because of this property, it is the only scale with which one can say object X is twice as large as object Y or three times as small as Y. Thus, this scale measures meaningful distance between objects. Although this scale has a solid theoretical foundation, it has some drawbacks in practice. First, it has been found difficult to use in field situations such as in consumer testing (Lawless

and Malone 1986a,b). Second, its application to complex stimuli has been questioned (Giovanni and Pangborn 1983). This scale is exemplified by the magnitude estimation procedure (Stevens and Galanter 1957) popularized in industrial applications by Moskowitz and associates (Moskowitz and Sidel 1971; Moskowitz 1977).

Briefly, in magnitude estimation the task of the panelist is to evaluate a stimulus in comparison with another where one of them is assigned a number, for example, 10. If the stimulus to be evaluated is three times as large, then the panelist gives a score of 30; if it is half as large, it is given a score of 5. Detailed description of the magnitude estimation experimental procedure is described by Moskowitz (1977).

Table 7.1-1 summarizes the statistical properties of the four classes of scales as a guide for practitioners in the selection of valid statistical operations. Recent studies showed that both ratio and category scales have similar sensitivity to finding perceptual differences in various types of consumer products (McDaniel and Sawyer 1981a,b; Pearce et al. 1986; Lawless and Malone 1986a,b).

7.2 THE THURSTONE-MOSTELLER MODEL

The Thurstone-Mosteller model is an indirect method of scaling where the scale values are derived from responses elicited by the stimuli. This method is useful for scaling large number of stimuli. In addition to obtaining a scale for each stimulus, a multiple comparison to determine significant differences between pairs of stimuli can be obtained. The Thurstone-Mosteller model is based on Thurstone's Law of Comparative Judgment (Thurstone 1927; Thurstone and Jones 1957), which equates the unknown psychological sensation to the observed frequency of a stimulus. As we shall see, the unit of measurement of Thurstone's law is the standard deviation.

Let P_{12} be the observed proportion by which stimulus R_1 is judged greater than stimulus R_2. Also let Z_{12} denote the value of the standard normal Z such that

$$P(Z > Z_{12}) = P_{12} \tag{7.2-1}$$

Denote by S_1 and S_2 the mean sensations evoked by stimuli R_1 and R_2, respectively. If π_{12} denotes the correlation between the two sensation distributions generated by stimuli R_1 and R_2, then Thurstone's law of comparative judgment is given by

$$S_1 - S_2 = Z_{12}\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho_{12}\sigma_1\sigma_2} \tag{7.2-2}$$

where σ_1^2 and $\sigma^2{}_2$ are, respectively, the variance of S_1 and S_2. The sensation distance $(S_1 - S_2)$ is assumed to correspond to the distance between R_1 and R_2. In Eq. (7.2-2) the variance cannot be estimated in advance; hence some simplifications are needed that leads to Thurstone's Case V of the Law of Comparative Judgment. Case V

assumes zero correlation and standard deviation equal to 1.0. Thus Eq. (7.2-2) becomes

$$S_1 - S_2 = Z_{12} \tag{7.2-3}$$

However, Mosteller (1951a,b) showed that one can obtain Eq. (7.2-3) by assuming equal standard deviation and correlation between sensations of any two stimuli. As clearly given by Eq. (7.2-3), the normal deviate Z_{12} is the distance between the unknown sensations S_1 and S_2.

In the form of Eq. (7.2-3), one can scale sensation by the comparison of a stimulus i with stimulus j using the paired comparison design. What is in effect perceived in this comparison is the order of sensations X_i and X_j. The amount of sensation X_i is assumed to be governed by the normal probability model with mean sensation S_i. The probability that sensation X_i is greater than sensation X_j is given by

$$P_{ij} = P(X_i > X_j) = \Phi(S_i - S_j) \tag{7.2-4}$$

where Φ denotes the standard normal cumulative distribution function given in Appendix Table J. Using this table, and denoting Z_{ij} as the (P_{ij})th percentile of the standard normal distribution, we can write Eq. (7.2-4) in the form

$$P_{ij} = \Phi(Z_{ij}) \tag{7.2-5}$$

noting that $Z_{ij} = S_i - S_j, \quad i = j = 1, 2, .., t$

from Eq. (7.2-3). For t = 3 stimuli, we have the following scale separation from a paired comparison experiment:

$$Z_{12} = S_1 - S_2$$
$$Z_{13} = S_1 - S_3$$
$$Z_{23} = S_2 - S_3$$

Finally, the estimates of mean scale sensation for the ith stimuli from the above paired comparison is

$$S_i = \Sigma Z_{ij}/t \tag{7.2-6}$$

Table 7.2-1 shows the layout of a paired comparison experiment showing the proportions P_{ij} and their respective normal deviates Z_{ij}. If a stimulus is compared to itself, the expected value of $P_{ii} = 0.50$, hence the corresponding $Z_{ii} = 0$ from Table J.

Table 7.2-1
Layout of paired comparison experiment showing proportions and normal deviates.

	i, i > j				
j	1	2	3	...	t
			Proportion		
1	P_{11}	P_{12}	P_{13}	...	P_{1t}
2	P_{21}	P_{22}	P_{23}	...	P_{2t}
...	...	...	...	...	...
...	...	...	...	...	...
t	P_{t1}	P_{t2}	P_{t3}	...	P_{tt}
			Normal deviate		
1	Z_{11}	Z_{12}	Z_{13}	...	Z_{1t}
2	Z_{21}	Z_{22}	Z_{23}	...	Z_{2t}
...	...	...	...	...	...
...	...	...	...	...	...
t	Z_{t1}	Z_{t2}	Z_{t3}	...	Z_{tt}
Sum	Z_1	Z_2	Z_3	...	Z_t
S_i	Z_1/t	Z_2/t	Z_3/t	...	Z_t/t

Example 7.2-1

This example illustrates the scaling of stimuli by the Thurstone-Mosteller model. Table 7.2-2 contains data on lather characteristics of a personal care product from a paired comparison experiment. In this experiment, 40 panelists were used to evaluate three treatments denoted by 1, 2, and 3. These treatments were assigned a 3-digit code number when they were evaluated. Each panelist evaluated all possible pairs of treatments. These pairs were 1 vs. 2, 1 vs. 3, and 2 vs. 3. The order of treatment presentation was done at random. In each pair, the panelists indicated which member of the pair has the most lather. The results show that 33 panelists indicated that treatment 1 has more lather than 2, 18 indicated that treatment 1 has more lather than 3, and finally 10 indicated that treatment 2 has more lather than 3. Based on these results, the estimates of the proportions P_{ij} are as follows:

$$P_{12} = 33/40 = 0.825,$$

$$P_{13} = 18/40 = 0.450,$$

$$P_{23} = 10/40 = 0.250.$$

It is always true that

Table 7.2-2
Paired comparison data (P_{ij}) on lather characteristic of personal care product; t = 3, n = 40.

	i, i > j		
j	1	2	3
		Proportion P_{ij}	
1	0	0.175	0.550
2	0.825	0	0.750
3	0.450	0.250	0
		Normal deviate Z_{ij}	
1	0	−0.935	0.126
2	0.935	0	0.674
3	−0.126	−0.674	0
Sum	0.809	−1.609	0.800
S_i	0.270	−0.536	0.267

$$P_{21} = 1 - P_{12} = 0.175$$

$$P_{31} = 1 - P_{13} = 0.550$$

$$P_{32} = 1 - P_{23} = 0.750$$

The above results are collected in Table 7.2-2.

The next step in the analysis is to transform the P_{ij}s to normal deviates Z_{ij}s by the aid of Table J. For example, we find that $P_{12} = 0.825$ or 82.5% corresponds to $Z_{12} = 0.935$, and so on. The value 0.935 is the distance between treatments 1 and 2 in standard deviation unit. In Table J, when P_{ij} is less than 50.0%, use $1 - P_{ij}$ as the proportion to find Z_{ij} with the sign reversed. For example, to find Z_{21} from $P_{21} = 0.175$, use $1 - 0.175 = 0.825$ which, from Table J is found to be -0.935 with sign reversed. Except for the sign, notice the symmetry of the estimates of Z_{ij} in Table 7.2-2.

The final step is to add algebraically each column of normal deviates and divide the sum by the number of treatments t to obtain the estimates of scale value S_i. For example,

$$S_1 = (0.935 - 0.126)/3 = 0.270$$

estimates the scale value for treatment 1. A negative value denotes a lesser amount of lather than a positive value. One may express scale values as a deviation from one of the treatments. If there is a control sample in the study, the control can be used as an anchor point on the scale continuum. Assuming that treatment 2 is the control sample, then

$$S_1 = 0.270 - (-0.536) = 0.806$$

$$S_2 = -0.536 - (-0.536) = 0.000$$

$$S_3 = 0.267 - (-0.536) = 0.803$$

In this form, the estimated scale values represent the scale separation between treatments which may be plotted as shown in Fig. 7.2-1. The significance of pairwise comparisons of treatments may be tested following the procedure given in Gacula and Singh (1984).

7.3 RANKING METHOD

This section discusses a ranking method of scaling developed by Dunn-Rankin (1965). The Dunn-Rankin method also provides tests of significance of the difference between scale values. Although there are other paired comparison methods of scaling such as the Guilford (1954) and the Bradley-Terry (1952) models, the simplicity of ranking procedure generally provides greater appeal to the user.

The experimental design for obtaining scaling data consists of ranking the stimuli from 1 to t or a paired comparison between stimuli totalling $t(t-1)/2$ comparisons. The steps in scaling by the ranking method are as follows:

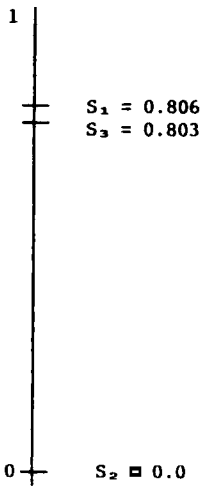


Fig. 7.2-1
Scale separation among treatments derived from the Thurstone-Mosteller model.

1. Determine the rank order of stimuli by complete ranking or ranking obtained by the method of paired comparison. If the stimuli are ranked in order of preference, for example, then values 1, 2, .., t are assigned to the stimuli. In preference studies, a stimulus with a rank of 1 would be the most preferred, whereas a rank of t would be the least preferred. If the ranking is obtained from a paired comparison, a frequency matrix is constructed containing the number of times that the ith stimulus is preferred over the jth stimulus. On a given pair, a value of 1 is assigned to the preferred stimulus and a 2 otherwise.

2. Sum the ranks for each stimulus over all n panelists, and add the total number of panelists to each rank sum. A check for the rank sum R_i, i = 1, 2, .., t, is found by the formula

$$R_i = n(t)(t + 1)/2 \tag{7.3-1}$$

3. Determine the minimum R(min) and maximum R(max) rank sums as well as the average rank $\overline{R}$ as follows:

$$\overline{R} = n(t + 1)/2 \tag{7.3-2}$$

$$R(\text{min}) = n, \quad R(\text{max}) = n(t) \tag{7.3-3}$$

4. Determine the rank sum standard deviation S by

$$S = \sqrt{n(t)(t + 1)/12} \tag{7.3-4}$$

5. Determine the normalize scale Z_i for each stimulus by

$$Z_i = (R_i - \overline{R})/S \tag{7.3-5}$$

where i runs from R(min) to R(max). The normalized scale is later converted to another scale so that it ranges from 0 to 100 by adding the absolute value of Z(min) to Z_i, then dividing each result by the largest new value, and multiplying by 100.

A multiple comparison of the estimated scales can be made using the original rank sum for each stimulus as described in Section 1.6 of Chapter 1.

Example 7.3-1

Let us use the data in Example 7.2-1 to illustrate the ranking method of scaling. Table 7.3-1 shows the data in the form of frequency matrix. Note that the number of panelists (n = 40) is added to the rank sum to obtain R_i. Relevant statistics necessary to complete the scaling are as follows:

$$\Sigma R_i = 40(3)(3 + 1)/2 = 240$$

$$R(\text{min}) = 40$$

Table 7.3-1
Frequency matrix derived from paired comparison experiment; t = 3, n = 40.

		Frequency i, i > j			
		1	2	3	
	1	...	7	22	
j	2	33	...	30	
	3	18	10	...	
Rank sum		51	17	52	
n		40	40	40	
Rank R_i		91	57	92	$\Sigma R_i = 240$

Note: i > j denotes that the ith treatment has more lather than the jth treatment in a paired comparison.

$$R(\max) = 3(40) = 120$$

$$\overline{R} = 40(3 + 1)/2 = 80$$

$$S = \sqrt{40(3)(3 + 1)/12} = 6.32$$

Substituting the estimates of R_i into Eq. (7.3-5), the normalized scale Z_i is obtained.

$$R(\min) = (40 - 80)/6.32 = -6.33$$

$$\text{Treatment } 1 = (91 - 80)/6.32 = 1.74$$

$$\text{Treatment } 2 = (57 - 80)/6.32 = -3.64$$

$$\text{Treatment } 3 = (92 - 80)/6.32 = 1.90$$

$$R(\max) = (120 - 80)/6.32 = 6.33$$

From step 5, the absolute value of R(min) is 6.33, hence the largest Z_i is 6.33 + R(max) = 6.33 + 6.33 = 12.66. Finally, the estimates of scale values for treatments 1, 2, and 3 are

$$\text{Treatment } 1 = [(6.33 + 1.74)/12.66]100 = 63.7,$$

$$\text{Treatment } 2 = [(6.33 - 3.64)/12.66]100 = 21.2,$$

$$\text{Treatment } 3 = [(6.33 + 1.90)/12.66]100 = 65.0.$$

Note that the lower and the upper limits of the scale are 0 and 100, respectively. The result may be presented on a linear scale continuum shown in Fig. 7.3-1. Clearly, treatments 1 and 3 have the most lather.

The next step in the statistical analysis is the multiple comparison of treatments.

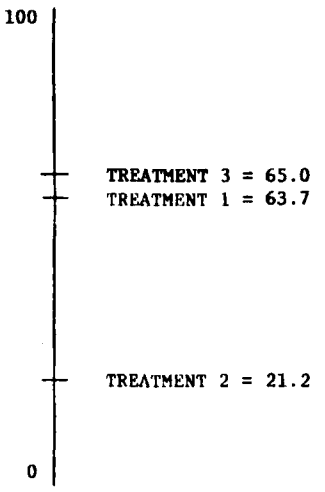


Fig. 7.3-1
Estimates of scale value for three treatments obtained by ranking method.

Remember that the original rank sum for each treatment is to be used. In this example, the absolute value of the pairwise differences are as follows:

$$d_{12} = 51 - 17 = 34$$

$$d_{13} = 51 - 52 = 1$$

$$d_{23} = 17 - 52 = 35$$

By consulting the critical range table (Appendix Table C) for $n = 40$, $t = 3$, and significance level of 1%, a critical range of 26 is found. This range should be equal to or exceeded by d_{ij} to declare significance. Since d_{12} and d_{23} exceeded the critical range of 26, these comparisons are significant.

Rank Scaling in Balanced Incomplete Block Designs

In the preceding discussion, the samples were completely compared to each other, i.e., AB, AC, BC, in pairs. In sensory studies, when the number of samples is large the number of pairwise comparisons becomes prohibitive and impractical for experimental execution. Therefore, one may use the balanced incomplete block designs discussed in Chapter 4. In this section, we will discuss scaling of stimuli by ranking in an incomplete block set-up, where only a k fraction of the total number of stimuli t is ranked by each panelist; hence, $k < t$. We will use the same definition of the parameters of balanced incomplete block design as discussed in Chapter 4.

The rationale of scaling by ranking is simple, in that if the stimuli are perceptibly differentiable then the stimuli can be ranked from 1 to t, and in the case of incomplete block design, from 1 to k, where k is the number of stimuli in each block, i.e., number of stimuli ranked by each panelist. Since incomplete block designs are balanced (See Chapter 4), the sum of ranks of the stimuli tends to be the same if the stimuli can not be differentiated. Durbin (1951) developed a statistic to test whether the rankings are the results of random assignment. The test statistic with $t - 1$ DF is defined by

$$T = \frac{12(t - 1)}{rt(k - 1)(k + 1)} \Sigma[R_i - r(k + 1)/2]^2 \qquad (7.3\text{-}5)$$

where R_i is the sum of ranks for the ith stimulus. Note that the letter r in Eq. (7.3-5) should be multiplied by p if the basic balanced incomplete block design is repeated p times. The null hypothesis of random assignment of ranks is rejected at an α-level of significance if T is equal to or greater than the Chi-square value obtained in Appendix Table I. An equivalent form of (7.3-5) is given by Skillings and Mack (1981).

In practice, if T is significant the scale values (sum of ranks) are subjected to a multiple comparison test. Skillings and Mack (1981) developed a multiple comparison test for large sample size. Letting $d_{ij} = \overline{R}_i - \overline{R}_j$, $i > j$, be the difference between the average rank sums, they showed that if

$$d_{ij} \geqslant q_{\alpha,t,\infty}[\sqrt{(k + 1)(rk - r + \lambda)/12r^2}] \qquad (7.3\text{-}6)$$

the difference between the average rank sums is significant at the α-level. The value of $q_{\alpha,t,\infty}$ is given in Appendix Table K for the 1 and 5% level of significance. Note that in Eq. (7.3-6), the parameters r and λ should be multiplied by p, the number of repetitions of the basic balanced incomplete block design.

Example 7.3-2

Ten prototypes were used to clean a clothing material, after which the treated materials were evaluated for overall fabric softness. Using a balanced incomplete block design shown in Table 7.3-2, each panelist ranked the clothing materials from 1 to 4 with 1 being soft and 4, not soft. The design in this table was repeated two times ($p = 2$) and the resulting data are shown in Table 7.3-3. The rank sum in this table is simply the sum of the ranks in each each column, and represents the scale value of each prototype as perceived by the trained 15-member panel.

To perform a multiple comparison of the scale values, the Durbin test is to be computed. Using Eq. (7.3-5),

$$T = \frac{12(10 - 1)}{12(10)(3)(5)}(460) = 27.60$$

Table 7.3-2
Fabric softness test using the balanced incomplete block design; t = 10, k = 4, r = 6, b = 15, λ = 2.

Panelist	Treatment	Order of Sample Presentation
1	1 2 3 4	2 4 1 3
2	1 2 5 6	5 1 6 2
3	1 3 7 8	1 7 3 8
4	1 4 9 10	4 9 1 10
5	1 5 7 9	9 5 1 7
6	1 6 8 10	8 6 10 1
7	2 3 6 9	2 6 3 9
8	2 4 7 10	4 2 10 7
9	2 5 8 10	10 8 5 2
10	2 7 8 9	2 7 9 8
11	3 5 9 10	3 9 5 10
12	3 6 7 10	10 7 3 6
13	3 4 5 8	5 4 8 3
14	4 5 6 7	4 7 5 6
15	4 6 8 9	6 4 8 9

Note: Design is taken from Cochran and Cox (1957).
Designs tabulated by Cochran and Cox should be randomized within panelists (within blocks).

which exceeds the Chi-square value of 21.666 (Appendix Table I, 9 DF) at the 1% level of significance. Therefore, the estimated scale values are not the result of random variation. The next step is the separation of the prototypes into groups using Eq. (7.3-6). Referring to Table K (Appendix) we obtain $q_{.05,10} = 1.86$. Therefore

$$d_{ij} \geqslant 1.86(\sqrt{[2(5)(12)(4) - 12 + 4)]/12(2)36)}$$

or

$$d_{ij} \geqslant 1.32,$$

where 1.32 is the critical value of the test. Any pairwise differences that are equal to or greater than 1.32 are declared significant at the 5% level. It is a matter of choice that the 5% level was used in the multiple comparison test, even though the Chi-square test was significant at the 1% level. The result of the grouping is given below.

Table 7.3-3
Rank data obtained by balanced incomplete block for Example 7.3-2.
Design parameters: t = 10, b = 15, r = 6, λ = 2, p = 2.

	Product									
Panelist	1	2	3	4	5	6	7	8	9	10
1	1	2	3	4						
2	1	3			2	4				
3	3		4				2	1		
4	4			3					1	2
5	2				3		1		4	
6	3					4		1		2
7		2	1			4			3	
8		2		4			3			1
9		1			3			4		2
10		3					2	4	1	
11			1		4				3	2
12			1			2	4			3
13			1	3	2			4		
14				4	2	3	1			
15				4		2		3	1	
16	3	1	2	4						
17	1	2			3	4				
18	2		4				3	1		
19	3			4					1	2
20	1				3		2		4	
21	3					4		2		1
22		1	3			4			2	
23		1		4			3			2
24		1			2			4		3
25		1					4	3	2	
26			1		3				2	4
27			1			2	4			3
28			1	3	2			4		
29				4	2	3	1			
30				4		1		2	3	
Rank sum	27	20	23	45	31	37	30	33	27	27
Average R_i	2.25	1.67	1.92	3.75	2.58	3.08	2.50	2.75	2.25	2.25

Note: p(r) = 2(6) = 12 replications per prototype.
p(λ) = 2(2) = 4 the number of times that a distinct pair appears in the design.

Prototype	Average rank	Grouping
2	1.67	a
3	1.92	a
1	2.25	ab
9	2.25	ab
10	2.25	ab
7	2.50	abc
5	2.58	abc
8	2.75	abc
6	3.08	bc
4	3.75	c

To illustrate, $R_8 - R_2 = 2.75 - 1.67 = 1.08$, which is less than 1.32 thus this difference is not significant. The grouping is interpreted as follows: Any two average ranks with one letter in common are not significantly different, i.e., 2.25 and 3.08 are not significantly different from one another, because they share a common letter "b," but 1.67 and 3.75 are significantly different.

7.4 TRANSITIVITY PROPERTY OF PAIRED COMPARISON

In a series of paired comparisons the correct choices of the underlying rank order depend on the physical distance between stimuli and the acuity of the sensory panels. If the perceivable difference between stimuli is well above threshold and the panels are adequately sensitive, the so-called transitivity property of paired comparison holds. This property is defined by the following:

$$\text{If stimulus A} > \text{B and B} > \text{C, therefore A} > \text{C.}$$

A deviation from transitivity implies that some pairs of stimuli are not distinguishable. The deviation from transitivity is called circular triads in the literature (David 1963). Circular triad occurs when

$$\text{C} > \text{A instead of A} > \text{C}$$

This pattern of judgments indicate that the stimuli are not scalable, therefore linear ordering is not possible. In practice this implies that the stimuli came from the same population.

In hypothesis testing, circular triads provide the sampling error for test of significance. Hence, sufficient replication should be used in an experiment to provide accurate estimate of this error. Circular triads can also be used as a measure

of inconsistency in the responses of judges/panelist (Slater 1961; Gridgeman 1963; Bauer 1978).

7.5 SCALING CONSUMER ACCEPTANCE

As a scientific discipline consumer acceptance pertains to the processes involved in human selection. Selection is generally based on past and present experiences of events or situations. Selection is not deterministic, but it is a temporal process (stochastic process) influenced by various personal and environmental factors, i.e., age, health, and needs. These factors are one of the reasons that consumer acceptance of products or services is difficult to measure at one time point. In designing a consumer acceptance study, it is very important that the appropriate experimental design should be used in order to obtain meaningful results.

Consumer acceptance can be scaled or measured in relation to other objects or products such as in preference testing, i.e.: "Which sample do you prefer?" It can also be measured in relation to consumer's frame of reference, such as in monadic testing. In other experimental situations, the scaling involves multiple sample comparisons.

Depending on the stage of product development and the intended use of the data, consumer acceptance can be measured in a laboratory setting (In-house consumer test), home-use setting (Home-use test), or in a central location test (CLT) setting. Let us briefly discuss these test procedures. For an in depth discussion of this area, one may refer to several marketing and sensory research publications (Amerine *et al.* 1965; Ferber 1974; Stone and Sidel 1985; Meilgaard *et al.* 1987; ASTM 1979).

In-House Consumer Test

Measurement of consumer acceptance under laboratory conditions is the cheapest way of procuring acceptance information. The in-house consumer test is done by consumers coming into the company's technical center or R&D laboratory to evaluate the product. Nally (1987) has discussed the establishment of consumer panels within laboratory/technical center. This paper should be consulted for those starting to develop a consumer panel data base.

The in-house test is more controlled and limited in scope. It is more controlled because the products to be evaluated are subjected to standardized preparation procedures. The advantage of this situation is lower variability in the data due to standardization, hence, a good estimate of product differences can be obtained. However, this estimate has built-in bias because of the panelists' knowledge of the laboratory. Although the samples are appropriately coded, panelists tend to avoid giving low or high ratings. Unfortunately, the magnitude of this bias is not known.

In-house consumer tests are limited to the evaluation of food, beverage, and other similar products where the consumer response is limited to taste, flavor, and ap-

pearance characteristics. Obviously, in-house testing is not suitable for evaluation of personal care products because panelists must use the products at home.

Home-Use Test

In the home-use test, the panelists either pick up the products at the R&D facilities or products are sent by a carrier for delivery to panelists' homes or other designated pick-up locations. It is recommended that only two products are to be evaluated by panelists in the home-use test in order to obtain reliable responses. Confusion, mistakes, and other logistic problems will result if many samples are evaluated. The use of incomplete block designs discussed in Chapter 4 are extremely useful in this situation.

The main advantage of this test procedure is that the products are used under normal home conditions, and, therefore, the test is more realistic. Uncontrollable factors during product use are built into the responses of the consumers, resulting in products that are robust. The subject of product robustness is discussed in Chapter 8. In addition, home-use testing is generally cheaper to conduct.

Central Location Test

This test is the most expensive to conduct, and generally is used at the later part of product development, i.e., prototypes have been finalized for testing against competitors or target brand category. However, there is evidence that CLT can be used early in product development thereby shortening the development process (Griffin and Stauffer 1991). Therefore, descriptive profiling and sensory difference testings may be bypassed and can proceed directly to product optimization by experimental design as discussed in Chapter 8. Although the initial investment is high by this method, the overall cost would be lower than the more traditional testing procedures.

Contract vendors conduct the CLT for a company. Briefly, the test is conducted in several locations, i.e., shopping malls, to obtain the appropriate demographics required by the test objectives. In addition to sensory questions, marketing questions are included in the questionnaire, i.e., price, purchase intent, and other demographic data.

Questionnaire Design

An important part of consumer testing is the development of the questionnaire used to gather responses. Three important issues must be addressed: type of scale, length of the questionnaire, i.e., how many attributes to be included, and the form of the questionnaire, which is tied to the experimental design of the proposed study. The hedonic scale (like/dislike) is the most widely used in consumer testing. In practice, the number of categories on this scale have varied from 7 to 9, i.e., 1 = dislike extremely, 5 = neither like nor dislike, 9 = like extremely. Another popular scale

for intensity rating is the 5-point "just right scale," i.e., 1 = much too weak, 3 = just right, 5 = much too strong. See Chapter 8 for the use of these scales in optimization of products for robustness.

In this section, we focus on questionnaire design for obtaining consumer responses. Figure 7.5-1 shows one form of questionnaire for paired comparison. The panelist evaluates the product side-by-side and the questionnaire provides a direct comparison of products A and B. The ratings are relative to each other, hence, this design would be sensitive to detecting sensory differences. However, it has been argued that this design exaggerates differences. The statistical analysis of the data from this design is the paired t-test, with the assumption that the differences are normally distributed. This assumption is usually satisfied in practice.

To counteract the argument that differences are exaggerated, one practice is to use the sequential monadic questionnaire design (Fig. 7.5-2). Each sample is evaluated singly as shown in this figure, with the assumption that panelist can provide absolute judgment. Here, the frame of reference during evaluation is internal to the panelist. However, since the evaluation is adjacent in time the ratings have the property of a paired comparison, i.e., scores are relative to each other, therefore the data are analyzed by the paired t-test. To simplify the logistic of the test, the sequential monadic often uses the paired comparison questionnaire (Fig. 7.5-3).

Another design is the pure monadic as shown in Fig. 7.5-4. Each sample is completely evaluated by different sets of panelists. The resulting data are analyzed by

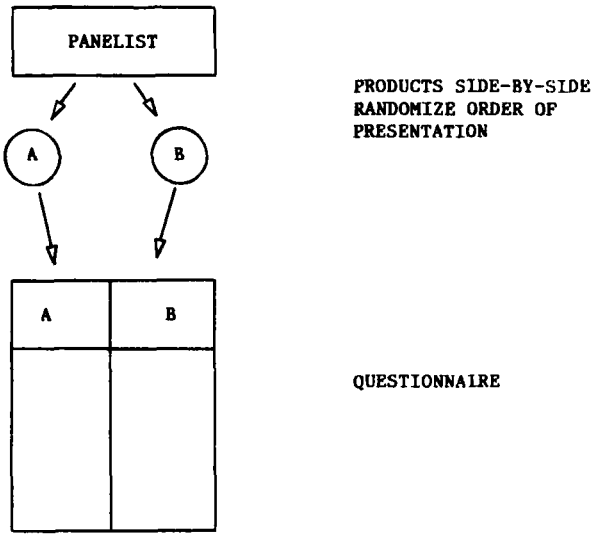


Fig. 7.5-1
Questionnaire for a paired comparison design. Samples and questionnaire form side-by-side.

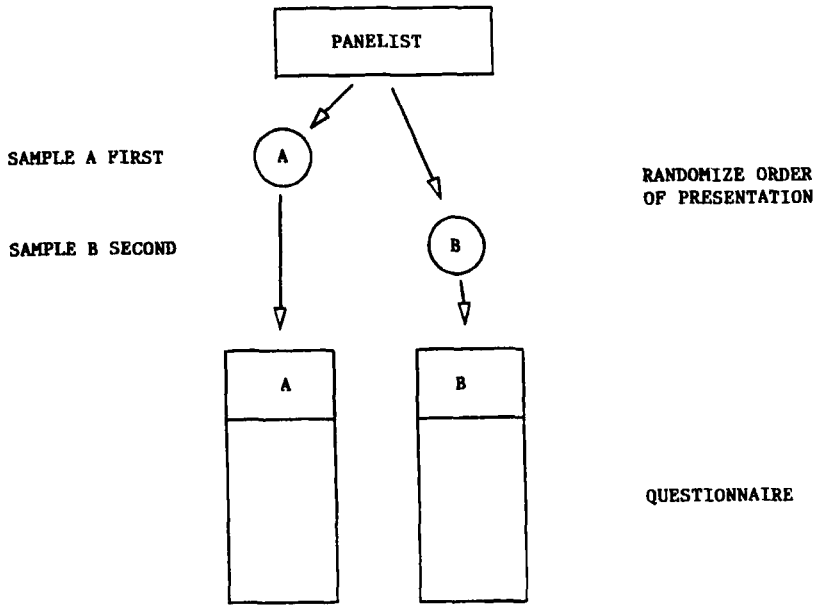


Fig. 7.5-2
Questionnaire for a sequential monadic design. Separate questionnaire form for each sample.

Table 7.5-1
Mean hedonic scores of *sultana* fruit evaluated monadically and by paired comparison (side-by-side) using a 9-point scale.

	Monadic design			Paired comparison		
Attribute	A	B	Difference	A	B	Difference
Initial test:						
Color	7.1	7.1	0.0	6.6	7.5	−0.9*
Flavor	7.1	7.2	−0.1	6.6	6.9	−0.3
Texture	7.1	7.4	−0.3	6.6	7.0	−0.4
Acceptability	7.0	7.3	−0.3	6.6	7.1	−0.5*
12-months from initial:						
Color	4.8	4.8	0.0	4.5	6.3	−1.8*
Flavor	5.4	5.9	−0.5	5.3	6.4	−1.1*
Texture	4.8	5.6	−0.8	5.9	6.4	−0.5*
Acceptability	4.8	5.5	−0.7	4.7	6.3	−1.6*

* Difference between means significant at the 5% level. Completely different panels were used to evaluate samples A and B in the monadic design. Source: McBride (1986).

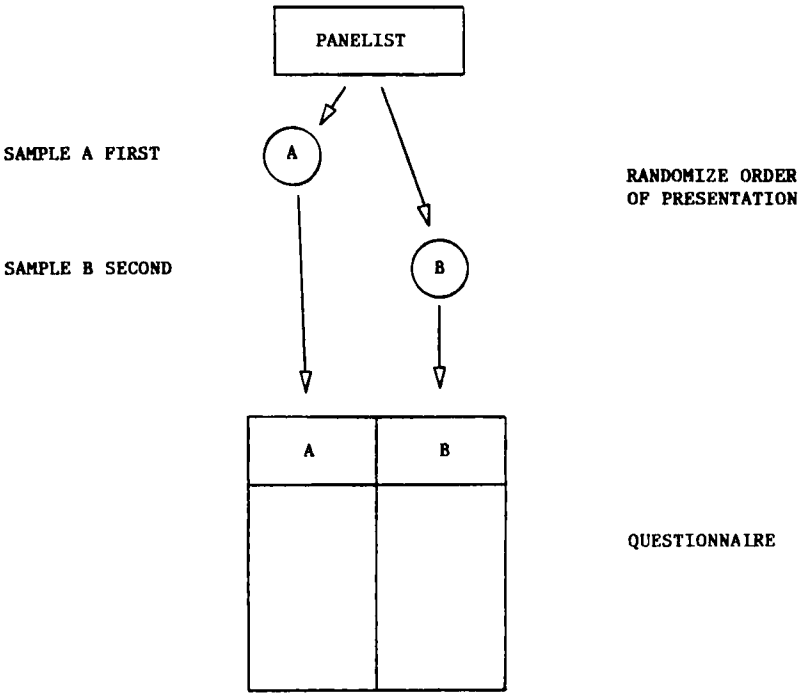


Fig. 7.5-3
Questionnaire for a sequential monadic design with paired comparison questionnaire form.

Table 7.5-2
Mean scores of samples A and B to illustrate that consumer sensory ratings are not absolute.

	Monadic	Paired comparison		
Attribute	A	A	B	Difference
Fragrance of bar	6.8	6.4	6.1	0.3
Strength of fragrance, %	65.4	55.6	51.5	4.1
Lather amount, %	79.8	65.7	68.4	−2.7
Moisturizing your skin	6.4	6.2	6.2	0.0
Leaving skin soft and smooth	7.0	6.4	6.4	0.0
Mildness	7.2	6.1	6.6	−0.5*

Note: Strength of fragrance and lather amount based on the 5-point just right scale; values reported are percentage of panelists falling on the just right category (category 3). The remaining attributes based on the 9-point hedonic scale.
* Significant at the 5% level.

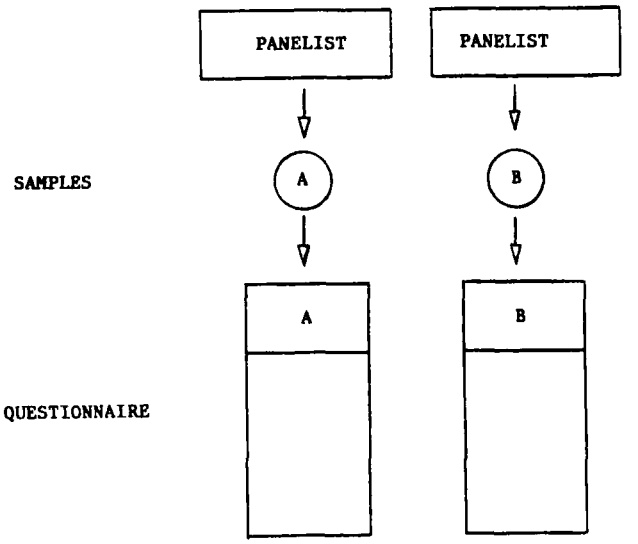


Fig. 7.5-4
Questionnaire for a pure monadic design. Samples and questionnaire form are independently completed.

the independent t-test. Here, the scores are relative to the frame of reference of the panelist. There are several assumptions that must be met to make the statistics valid. The first is that the ratings are normally distributed and this assumption is often violated. Second, the variance of samples A and B is approximately equal (common variance). Third, that consumer can reliably render absolute judgments, which is suspect. Furthermore, this design may require a larger number of panelists than the paired comparison design to counteract panelists variability. Note that in the analysis of paired comparison data the panelist variability is discounted by using the within panelist variance in the analysis.

Let us examine the data in Table 7.5-1 to illustrate the differences in the results obtained by the paired comparison and by the pure monadic design on the same sample. Notice that the differences between products favored sample B in both designs, but the paired comparison showed significant differences. Furthermore, the monadic design yielded higher mean scores than the paired comparison, indicating that consumer judgments are not absolute but are relative. This result is also supported by the data in Table 7.5-2. In general, when the products evaluated are highly acceptable there is a tendency for the monadic ratings to be high and vice versa. This phenomenon is known as "regression effect." Thus in using the monadic design, one must be careful in interpreting the data, for it may provide a false and misleading acceptance values. In real life situations, although products are used singly, the long run selection of products do involve comparative judgments internal to the consumer.

PRODUCT OPTIMIZATION

In this chapter, the optimization of product ingredients with the ultimate goal of producing the best product under a given set of criteria is discussed. The optimization principle that will be given here also applies to other forms of applications, such as optimization of process controls, transportation routing, and the number of cash registers to be installed in a given store. The emphasis in this chapter is on graphical solution of optimization problems, particularly in studies where multiple attributes are measured.

Measurement of multiple attributes or responses are common in the food, beverage, household, and the personal care industries. For example, in evaluating a food product one looks at simultaneously many attributes such as flavor, texture, color, and possibly cost. These attributes may carry different weights in the perception of the product to make it acceptable. How should one combine these attributes in order to maximize the acceptability of a product? A procedure for combining these attributes to provide a single measure of acceptability will be given. Some useful experimental designs will be discussed and illustrated with examples. Readers may refer to Box and Draper (1987), Gacula and Singh (1984), and to Myers (1971) for the statistical aspects of optimization methods.

8.1 PRELIMINARIES

The word optimization has several meanings depending on the context in which the word is used. In general, optimization is a series of steps for obtaining the best result under a given set of circumstances. Specific definitions relating to product formulation have been reported (Sidel and Stone 1983; Moskowitz 1983; Fishken 1983), and all definitions similarly led to producing the best product in its class. In this section, we give a mathematical definition of optimization as the process of finding the conditions that give the maximum or minimum value of a function. Once this function is obtained, one can find the best product formulation in its class.

The equation

$$Y = B_0 + B_1X_1$$

is an example of a function. In this equation, Y is known as the dependent variable, such as product attributes, X_1 is the independent variable, such as product ingre-

dients, and B_0 and B_1 are the parameters of the equation to be estimated from the experimental data. Note that Y can be a function of several independent variables. For example

$$Y = B_0 + B_1X_1 + B_2X_2$$

Here, Y is a function of variables X_1 and X_2. Figures 8.1-1 and 8.1-2 are graphical examples of several functions (Rao 1984). In these figures, the values of Y vary with the values of X. The relative maximum and minimum play an important role in the effective search for the best product formulation. In some stages of product development the services of expert/consultant can simplify the search for the operating window where optimization studies can start.

Test for Adequacy of Statistical Model

The equation that we have discussed is also called a statistical model that describes the relationship between the dependent variable Y and the independent variable X. How do we know that such a relationship is adequately described by the model? This section will deal with this question.

The most common models encountered in optimization studies to describe such a relationship are the so-called first-order and second-order models. The first-order model is given by a simple linear regression relationship

$$Y = B_0 + B_1X_1 + B_2X_2 + .. + B_kX_k + E \qquad (8.1\text{-}1)$$

whereas a second-order model is described by a quadratic regression relationship given by

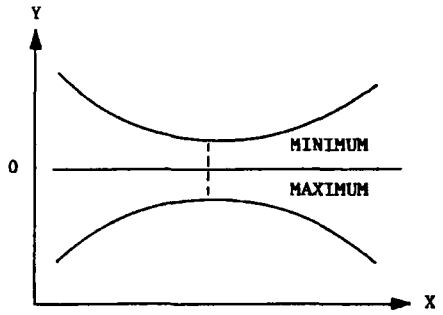


Fig. 8.1-1
Plots of maximum and minimum points of a function.

$$Y = B_0 + B_1X_1 + .. + B_kX_k + B_{11}X_1^2 + .. + B_{kk}X_k^2 + . . + B_{12}X_1X_2 + .. + B_{k-1,k}X_{k-1}X_k + E \quad (8.1\text{-}2)$$

where the parameters of the models are defined as follows:

B_0 = the intercept (grand mean), and its estimate is denoted by b_0.
B_i = the linear effect of X_i, and its estimate is denoted by b_i, i = 1, .., k.
B_{ii} = the quadratic effect of X_i, and its estimate is denoted by b_{ii}, i = 1, .., k.
B_{ij} = the interaction effect of X_i and X_j, and its estimate denoted by b_{ij}, i < j, i = 1, .., k−1 and j = 1, .., k.
E = random errors assumed to be normally distributed with mean 0 and variance σ^2. This variance is estimated by the error mean square (MSE).

The analysis of variance tables for the first-order and the second-order models are shown in Tables 8.1-1 and 8.1-2, respectively. As discussed in Chapter 6, the error (SSE) can be divided into two portions, one due to lack of fit (SSFIT) and the other due to experimental error (SSPURE). The lack of fit measures the inadequacy of the fitted model, while the experimental error, also known as pure error, measures the random error in the data obtained from replicated design points.

The design for fitting the first order model is the 2^k factorial design shown in Fig. 8.1-3 for a two-variable design. The fitting of first order model requires only 2 levels of each variable denoted by high (1) and low (−1). The SSPURE can be obtained by the addition of a center point to the design shown in Fig. 8.1-3 appropriately replicated. The design with the center point is shown in Fig. 8.1-4. In addition, if the factorial design points are also replicated these points can be included in the estimation of SSPURE.

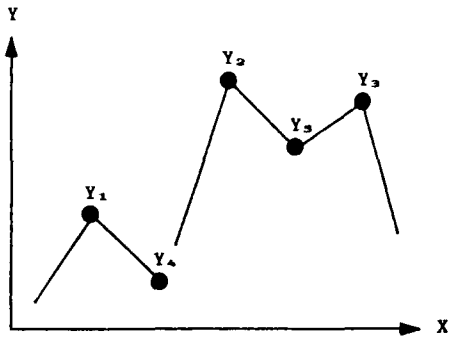


Fig. 8.1-2
Plots of relative and global points of a function: Y_1, Y_2, Y_3 = relative maxima; Y_2 = global maximum; Y_4, Y_5 = relative minima, Y_4 = global minimum.

Table 8.1-1
Analysis of variance for the first-order model.

Source of Variance	DF	SS	MS
Total	M-1	SSTO	
Linear regression:	k	SSR	MSR
B_1	1		
B_2	1		
•	•		
•	•		
B_k	1		
Error:	M-k-1	SSE	MSE
Lack of fit	N-k	SSFIT	MSFIT
Experimental error	M-N-1	SSPURE	MSPURE

Note: M = total number of observations.
N = number of factorial design points, excluding the center point.
k = number of factors or independent variables.
SSE = SSTO − SSR
SSFIT = SSE − SSPURE

To test for the adequacy of the fitted model the F-ratio is MSFIT/MSPURE. If the F-ratio is significant it indicates that the first-order model is unsatisfactory. Then the model should be revised to include interaction effects and the model becomes a second-order without the quadratic effects.

The design for the second-order model must have at least 3 levels of each variable to be able to estimate the quadratics plus the interaction and linear effects. Figure 8.1-5 is an example of a design for the second-order model, which is a 3^2 factorial design as shown in Chapter 6. As will be discussed later, if one rotates this design we obtain the central composite design, which is very useful in optimization work.

Least Squares Estimation of Regression Parameters

The regression parameters and their corresponding sums of squares (SS) shown in Tables 8.1-1 and 8.1-2 can be estimated by the method of least squares. It is convenient to present the least squares normal equations in matrix form given by

$$X'Y = (X'X)B \tag{8.1-3}$$

where X is the design matrix and Y is a column vector of responses. The design matrix can be the actual levels used in the experiment or the coded levels, i.e., 1, −1. The actual level can be coded by the formula

$$\text{Coded level} = (L - C)/S \tag{8.1-4}$$

Table 8.1-2
Analysis of variance for the second-order model.

Source of Variance	DF	SS	MS
Total	M-1	SSTO	
Linear regression:	k	SSR	MSR
B_1	1		
B_2	1		
•	•		
•	•		
B_k	1		
Quadratic:	k	SSQ	MSQ
B_{11}	1		
B_{22}	1		
•	•		
•	•		
B_{kk}	1		
Interaction:	k(k−1)/2	SSI	MSI
B_{12}	1		
B_{13}	1		
•	•		
•	•		
$B_{k-1,k}$	1		
Error:	Difference	SSE	MSE
Lack of fit	N-k	SSFIT	MSFIT
Experimental error	M-N-1	SSPURE	MSPURE

Note: M = total number of observations.
N = number of factorial design points, excluding the center point.
k = number of factors or independent variables.
SSE = SSSTO − SSR − SSQ − SSI
SSFIT = SSE − SSPURE

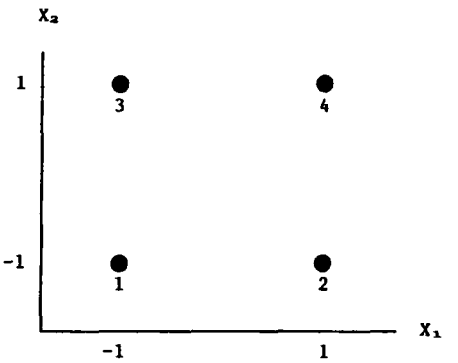


Fig. 8.1-3
Design points for a 2^2 factorial for fitting a first-order model.

where L is the level of interest, C is the middle level, and S is the width between the high and the low levels of interest equally spaced. Coded levels take on values of -1, 0, and 1 corresponding to the low, middle, and high levels.

For example, the matrix for the design in Fig. 8.1-3 with the constant column B_0 included is

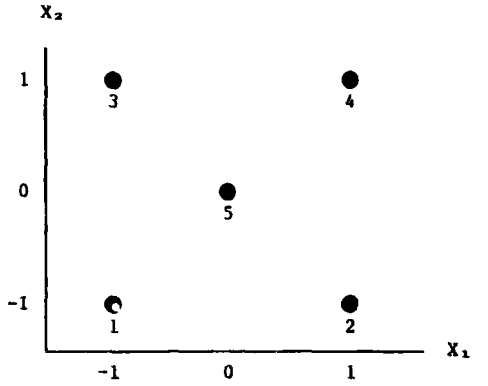


Fig. 8.1-4
Design points of a 2^2 factorial with center point.

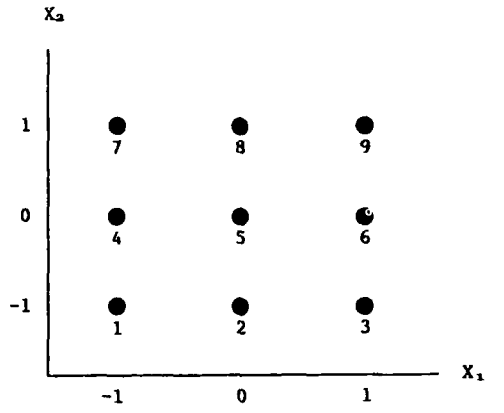


Fig. 8.1-5
Design points of a 3^2 factorial for fitting a second-order model.

$$X = \begin{array}{c} \begin{matrix} B_0 & X_1 & X_2 \end{matrix} \\ \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \end{array}$$

Notice that excluding B_0 the rows of matrix X represents the design points of Fig. 8.1-3. The transpose of matrix X is

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix}$$

where the columns of matrix X becomes rows of X'. The column vector of responses is

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix}$$

In most sensory work the responses are means of panelists or judges. Solving for B in formula (8.1-3) gives the least squares estimates

$$b = (X'X)^{-1}X'Y \tag{8.1-5}$$

provided that the design matrix has an inverse or a generalized inverse if necessary. It is also known that the variance (Var) of b is the diagonal elements of the variance-covariance matrix $\sigma^2(X'X)^{-1}$, where σ^2 is estimated by MSE in the analysis of variance. For the design in Fig. 8.1-3, $\text{var}(b_0) = \text{var}(b_1) = \text{var}(b_2) = \sigma^2/4$. Statistical software generally prints out the variance-covariance matrix.

In this book, we will use either SAS (SAS Institute), STATISTIX (NH Analytical Software), DESIGN-EXPERT (Stat-Ease, Inc.) or X-STAT (Wiley Professional Software) to evaluate formula (8.1-3). These commercial software packages require the design matrix as input to run the program.

Denoting the estimates of X'Y by g, then

$$\begin{aligned} SSR &= b_1 g_1 + b_2 g_2 + b_{12} g_{12} \\ SSI &= b_{12} g_{12} \\ SSQ &= b_{11} g_{11} + b_{22} g_{22} \end{aligned} \tag{8.1-6}$$

The above formulas can be extended to include more than two variables. Furthermore

$$SSTO = \Sigma Y_i^2 - [(\Sigma Y_i)^2/M], \quad i=1, .., M \tag{8.1-7}$$

$$SSPURE = \Sigma\{\Sigma Y_i^2 - [(\Sigma Y_i)^2/r]\}, \quad i=1, ..,r \tag{8.1-8}$$

where r is the number of replications per design point. Note that SSPURE is summed over the M design points.

If the design point is not replicated, instead the design is augmented such as that shown in Fig. 8.1-4 with center point replicated r times, then

$$SSPURE = \Sigma Y_i^2 - [(\Sigma Y_i)^2/r] \tag{8.1-9}$$

Further reading on lack of fit testing when replications are not available is given by Green (1971), Shillington (1979), and Joglekar *et al.* (1989). Let us illustrate the estimation of regression parameters by an example.

Example 8.1-1

Consider the four design points in Fig. 8.1-3 with observed values of each point to be 5.3, 6.0, 6.4, and 7.6, respectively. In matrix form, this is written as a column vector Y:

$$Y = \begin{bmatrix} 5.3 \\ 6.0 \\ 6.4 \\ 7.6 \end{bmatrix}$$

Likewise, the coordinates of design points 1, 2, 3, and 4 are $(-1,-1)$, $(1,-1)$, $(-1,1)$, and $(1,1)$, respectively. Using these coordinates, the design matrix with the B_0 column of 1s included is

$$X = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

and multiplying matrix X by its transpose X′, we obtain

$$X'X = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

Also

$$X'Y = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 5.3 \\ 6.0 \\ 6.4 \\ 7.6 \end{bmatrix}$$

$$= \begin{bmatrix} 25.3 \\ 1.9 \\ 2.7 \end{bmatrix}$$

Using Eq. (8.1-5), the least squares estimates of B is

$$b = \begin{bmatrix} .25 & 0 & 0 \\ 0 & .25 & 0 \\ 0 & 0 & .25 \end{bmatrix} \begin{bmatrix} 25.3 \\ 1.9 \\ 2.7 \end{bmatrix} = \begin{bmatrix} 6.325 \\ 0.475 \\ 0.675 \end{bmatrix}$$

hence, $Y = 6.325 + 0.475X_1 + 0.675X_2$.

The sum of squares due to regression is from Eq. (8.1-6)

$$SSR = 0.475(1.9) + 0.675(2.7) = 2.7250 \text{ with } k = 2 \text{ DF}$$

and from formula (8.1-7)

$$SSTO = 162.81 - [(25.3)^2/4] = 2.7875 \text{ with } M-1 = 4-1 = 3 \text{ DF.}$$

Also

$SSE = SSTO - SSR = 2.7875 - 2.7250 = 0.0625$ with $M-k-1 = 4-2-1 = 1$ DF. The ANOVA is given in Table 8.1-3. With a limited number of degrees of freedom, it is meaningless to compute the F ratio. Using the estimate of error one computes the variance-covariance matrix as follows:

Table 8.1-3
ANOVA for a first-order model with two variables (Example 8.1-1).

Source of Variance	DF	SS	MS
Total	3	2.7875	
Regression:	2	2.7250	
b_1	1	0.9025	
b_2	1	1.8225	
Error	1	0.0625	0.0625

Note: 2.7250 = 0.9025 + 1.8225

$$\sigma^2(X'X) = 0.0625 \begin{bmatrix} .25 & 0 & 0 \\ 0 & .25 & 0 \\ 0 & 0 & .25 \end{bmatrix} = \begin{bmatrix} 0.0156 \\ 0.0156 \\ 0.0156 \end{bmatrix}$$

hence, $\text{var}(b_0) = \text{var}(b_1) = \text{var}(b_2) = 0.0156$. Notice that the parameters are estimated with equal accuracy. The square root of these quantities provides the standard error of the regression coefficients, i.e., $\sqrt{0.0156} = 0.12$.

The various calculations that we have done serves as an insight into regression analysis and should also aid the reader in understanding the output of computer programs. For example, Table 8.1-4 shows the output of our example using STATISTIX. There are two things that should be pointed out in this table. The first is the Student's t statistic, which tests the hypothesis that the regression coefficient is equal to zero. The second is the R-square statistic, which indicates the goodness of fit of the model to the data. The R-square ranges from 0 to 1.0 with high values indicating a good fit. In this book, it is recommended that the adjusted R-square should be used because they are adjusted for the number of parameters in the model. It should be pointed out that R-square values can be increased by including more parameters in the model, and when blindly done can lead to misleading results. See also Section 8.3.

Tables 8.1-5 and 8.1-6 show the variance-covariance matrix and the ANOVA table, respectively. The reader may compare these results to the manual calculations used in this example.

Example 8.1-2

In this example, the design used in Example 8.1-1 is augmented by adding a center point replicated 5 times (r = 5) with the following observed values: 6.3, 6.0, 5.9, 6.1, and 6.4. Now one can estimate the pure error to test for the adequacy of the model. Using Eq. (8.1-9),

Table 8.1-4
STATISTIX output for Example 8.1-1.

```
STATISTIX 3.5
ID: TABLE 8.1-4

VIEW DATA

 CASE        X1          X2          Y

    1   -1.0000     -1.0000     5.3000
    2    1.0000     -1.0000     6.0000
    3   -1.0000      1.0000     6.4000
    4    1.0000      1.0000     7.6000


UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF Y

PREDICTOR
VARIABLES     COEFFICIENT     STD ERROR      STUDENT'S T       P
---------     -----------     ---------      -----------     ------
CONSTANT       6.3250         1.2500E-01       50.60        0.0126
X1             4.7500E-01     1.2500E-01        3.80        0.1638
X2             6.7500E-01     1.2500E-01        5.40        0.1166

R SQUARED              0.9776     RESID. MEAN SQUARE (MSE)   6.250E-02
ADJUSTED R SQUARED     0.9327     STANDARD DEVIATION         2.500E-01

SOURCE        DF       SS            MS           F        P
------        --       --            --           -        -
REGRESSION     2   2.7250        1.3625        21.80    0.1497
RESIDUAL       1   6.2500E-02    6.2500E-02
TOTAL          3   2.7875

CASES INCLUDED 4   MISSING CASES 0
```

Table 8.1-5
Variance-covariance matrix for Example 8.1-1. Note that the off-diagonal should be treated as zeros.

```
STATISTIX 3.5
ID: EX811

VARIANCE - COVARIANCE MATRIX FOR COEFFICIENTS

              CONSTANT     X1           X2
CONSTANT      1.562E-02
X1           -2.168E-19    1.562E-02
X2           -2.168E-19    8.674E-19    1.562E-02
```

$$\text{SSPURE} = (6.3^2 + .. + 6.4^2) - (30.7^2/5)$$
$$= 0.1720$$

Table 8.1-6
ANOVA table for Example 8.1-1 using the STATISTIX software.

```
STATISTIX 3.5                                              30 DEC 91, 22:18
ID: EX811

STEPWISE ANALYSIS OF VARIANCE OF Y

SOURCE      INDIVIDUAL SS   CUM DF   CUMULATIVE SS   CUMULATIVE MS   ADJUSTED R-SQUARED   MALLOW'S CP   P
CONSTANT    160.02
X1          9.0250E-01          1   9.0250E-01      9.0250E-01      -0.0143              30.2          2
X2          1.8225              2   2.7250          1.3625           0.9327               3.0          3
RESIDUAL    6.2500E-02          3   2.7875          9.2917E-01

R SQUARED                0.9776     RESID. MEAN SQUARE (MSE)   6.250E-02
ADJUSTED R SQUARED       0.9327     STANDARD DEVIATION         2.500E-01
```

Using the same procedure in Example 8.1-1, the following sums of squares are obtained:

$$\text{SSTO} = 3.0356, \qquad \text{SSR} = 2.7250, \qquad \text{SSE} = 0.3106$$

Then

$$\text{SSFIT} = 0.3106 - 0.1720 = 0.1386.$$

These sums of squares are collected in the ANOVA Table 8.1-7. Notice that we have partitioned the error sum of squares into two components. The F-ratio

$$F = 0.0693/0.0430 = 1.611$$

Table 8.1-7
ANOVA table for the augmented design used in Example 8.1-2.

Source of Variance	DF	SS	MS	F-ratio
Total	8	3.0356		
Regression:	2	2.7270		
b_1	1	0.9025	0.9025	17.42**
b_2	1	1.8225	1.8225	35.18**
Error:	6	0.3106	0.0518	
Lack of fit	2	0.1386	0.0693	1.61
Pure error	4	0.1720	0.0430	

**$p < .01$

with 2 numerator DF and 4 denominator DF is not significant, hence the first-order model appears adequate to describe the relationship. The R-square statistics are computed as follows:

$$\text{R-square} = 1 - (\text{SSE/SSTO}) = 1 - (0.3106/3.0356) = 0.8977$$

$$\text{Adjusted R-square} = 1 - [(M-1)\text{MSE/SSTO}] = 1 - [(9-1)0.0518/3.0356] = 0.8636$$

The linear regression analysis using STATISTIX is given in Table 8.1-8. As expected, the regression coefficients did not change from those obtained in Example

Table 8.1-8
STATISTIX output for Example 8.1-2.

```
STATISTIX 3.5
ID: TABLE 8.1-8

VIEW DATA

 CASE        X1          X2          Y

    1   -1.0000     -1.0000     5.3000
    2    1.0000     -1.0000     6.0000
    3   -1.0000      1.0000     6.4000
    4    1.0000      1.0000     7.6000
    5    0.0000      0.0000     6.3000
    6    0.0000      0.0000     6.0000
    7    0.0000      0.0000     5.9000
    8    0.0000      0.0000     6.1000
    9    0.0000      0.0000     6.4000

UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF Y

PREDICTOR
VARIABLES     COEFFICIENT     STD ERROR     STUDENT'S T       P
---------     -----------     ---------     -----------     ------
CONSTANT      6.2222          7.5836E-02       82.05        0.0000
X1            4.7500E-01      1.1375E-01        4.18        0.0058
X2            6.7500E-01      1.1375E-01        5.93        0.0010

R SQUARED             0.8977      RESID. MEAN SQUARE (MSE)    5.176E-02
ADJUSTED R SQUARED    0.8636      STANDARD DEVIATION          2.275E-01

SOURCE          DF     SS             MS            F         P
------          --   ----------    ----------     -----    ------
REGRESSION       2   2.7250        1.3625         26.32    0.0011
RESIDUAL         6   3.1056E-01    5.1759E-02
TOTAL            8   3.0356

CASES INCLUDED 9    MISSING CASES 0
```

8.1-1, since we did not add or delete any observations in points 1, 2, 3, and 4. However, the intercept changed as it is an estimate of the grand mean, which now includes the center point observations. With the increased number of design points, the variances of the regression parameters have decreased (Table 8.1-9):

$$\text{Var}(b_0) = 0.0058, \qquad \text{Var}(b_1) = \text{Var}(b_2) = 0.0129$$

8.2 WHY USE AN OPTIMIZATION TECHNIQUE?

There are several reasons why one should use an optimization technique in product formulation work. Not only is it a proven technique, its application has been greatly facilitated by the availability of computing technology in both hardware and software, which were the limiting factors in the past. The advantages of using an optimization technique are as follows:

1. It is fast and cost-effective by avoiding experimental reruns.
2. It is statistically efficient because the interaction among the independent variables and their quadratic effects on the response attributes (dependent variable) can be estimated simultaneously.
3. Optimization method provides a data base to answer direct questions such as What if: what if ingredient X becomes expensive and would like to reduce the amount in the formulation; what happens to the sensory and/or physical characteristics of the product? It is not necessary to run a full pledge experiment since the data base can provide the required information. However, a validation experiment is necessary.
4. Optimization method provides several potential product formulas for consumer evaluation.
5. Optimization method provides R&D direction to meet the changing market demands.
6. Optimization method provides discipline in the conduct of scientific research, and most importantly generates quality data.

Table 8.1-9
Variance-covariance matrix for Example 8.1-2. Note that the off-diagonal values should be treated as zeros.

VARIANCE - COVARIANCE MATRIX FOR COEFFICIENTS

	CONSTANT	X1	X2
CONSTANT	5.751E-03		
X1	-7.981E-20	1.294E-02	
X2	-7.981E-20	7.183E-19	1.294E-02

However, there are pitfalls in the use of optimization methods if one is not careful during the planning stage of the study. These pitfalls are the following:

1. Important factors/variables that affect the response have not been correctly identified. The investigator should know the function of each ingredient in the formulation or have some theory about their effects on the response.
2. The lower and upper levels of the factors have been incorrectly specified. Preliminary work is needed if these levels are not available. Some estimates may be obtained from existing products and such estimate would be the middle level in the design specification.
3. Over-use of extrapolation of response surface maps without appropriate check points.
4. Use of incorrect statistical model and experimental design.
5. Failure to recognize the presence of large systematic and/or random variation in the response.
6. Failure to recognize that the response function has more than one optimal point in the response surface.
7. Failure to verify the correctness of the selected optimum formulas against a control or standard formula.

The success in the use of optimization methods comes from experience and understanding of the biological, physical, or chemical system that one is working with. In using sensory measures to obtain responses, it should be emphasized that the selected optimum formulas should always be evaluated against a standard or control formulation before a final decision is to be made. This part of optimization is often overlooked.

The generalized steps in an optimization study are as follows:

1. Clearly define the objective of the study. This is perhaps one of the most important aspects in optimization work. Limit the objectives of the study, instead of being carried away to answering a multitude of objectives in the hope of saving experimental cost. This tactic usually results in futility.
2. Identify the variables to be studied. Historical data and information from published work are good sources. If they are not available, a preliminary bench work should be done.
3. Identify the response variables, that is, the attributes to be measured. Again, limit the attributes to be measured to those relevant to the study, perhaps not to exceed 10 attributes, to avoid confusion in sensory evaluation.
4. Set the lower and upper levels of the variables through preliminary experimentation or from prior knowledge.

5. Construct the appropriate experimental design. The design should be able to provide answers to the test objectives stated in Step 1.
6. Conduct the experiment following the design, and adjust the design strategy if needed. Collect and record the data with care.
7. Obtain the model for fitting each response variable and test for the adequacy of fit of the model to the data.
8. Construct contour maps and obtain optimum areas. Use the contour overlay technique for multi-response study. Review contour maps with the investigator to obtain feedbacks of areas where concentration should be directed.
9. Decide on potential formulations and make the product. Test 2 or 3 experimentally determined optimal formulations against a standard/control or against an established product brand.
10. Fine-tune optimal formulations, if necessary.

The number of steps in an optimization procedure can vary to a large degree (Fishken 1983; Sidel and Stone 1983; Giovanni 1983; Schutz 1983; Moskowitz 1983). As reviewed by LaGrange and Norback (1987), all of these reports have similarities which include: (1) an initial development study during which prototypes are developed and critical input variables are identified, (2) a screening/product recipe development step, which includes the determination of ingredient and processing levels of variables, and (3) a formal optimization study, with or without constraints, that includes consumer testing, data analysis, reformulation, and implementation.

As an overview of the various steps involved in the design and analysis of optimization studies, Fig. 8.2-1 is constructed to provide these steps at a glance. Note that fractional designs and the data collection designs were discussed in Chapters 2, 4, and 6. The Thurstone-Mosteller model of analysis was illustrated in Chapter 7.

8.3 TYPES OF OPTIMIZATION EXPERIMENTS

Before embarking an optimization experiment, it is important to know the nature of the ingredients (independent variables) and the response attributes to be measured. The reason for this is that the nature of the ingredients and the responses determines the type of optimization design.

First, determine whether the response depends on the proportion or the amount of the ingredient in the formula. For example, will the taste be the same for a 50:50 mixture of sugar and water in a glass as compared to the same mixture in a gallon container? If it is the same, then we have the so-called mixture experiment. If they are not, we call this type of study a nonmixture experiment. Let us refer to the configuration in Fig. 8.3-1, and call X_1 the amount of water and X_2 the amount of sugar in the mixture. Assume we have the following information:

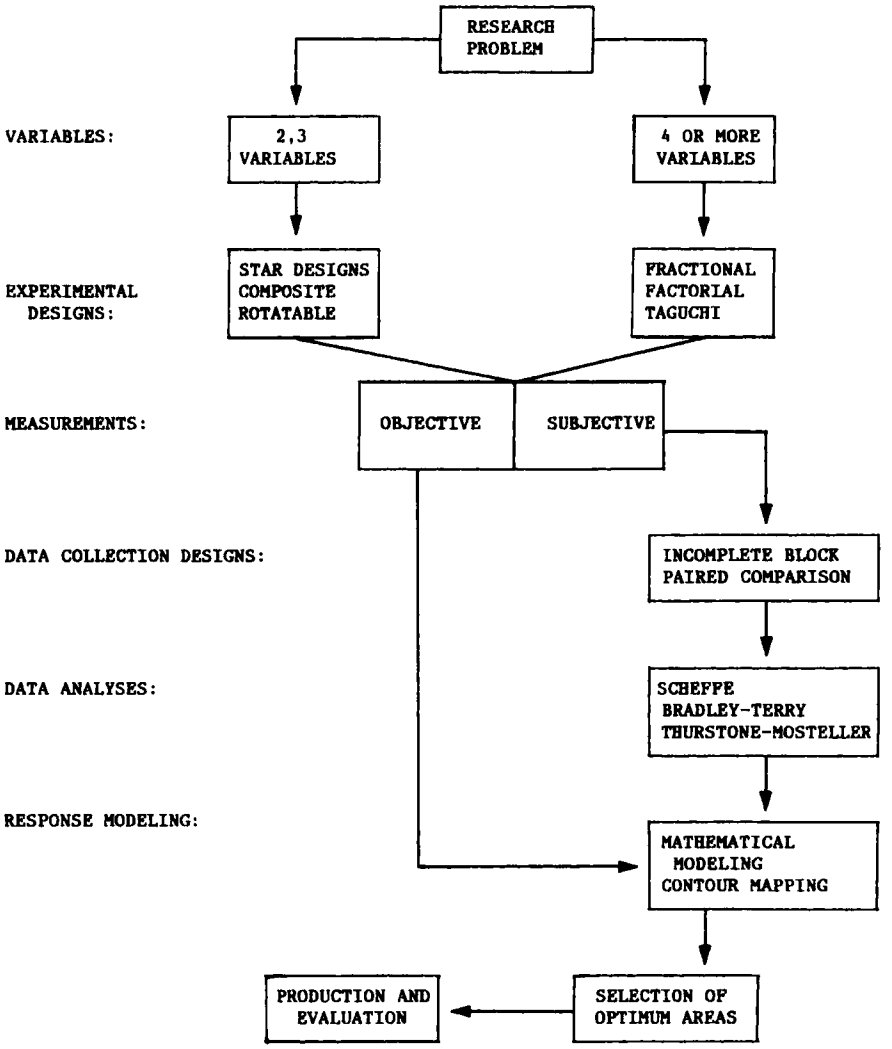


Fig. 8.2-1
Steps in design and analysis of optimization studies.

Design pt.	Water, gr.	Sugar, gr.
1	6	2
2	12	2
3	6	4
4	12	4

Notice that design points 1 and 4 have the same proportion of water and sugar, which is 33.3% sugar. If the response depends only on the proportion of the ingredients in the mixture, then we do not need design point 4. The consequence of this result is discussed below.

Nonmixture Experiments

In this type of experiment the levels of the variables (factors) are unrestricted and independent of each other. That is, changing the in-going levels of one variable does not affect the levels of the other variables in the formulation. The response attribute in nonmixture experiments depends only on the absolute amount used in the formulation. For example, one may have a formulation as follows:

Ingredient	Amount, lb.
A	0.5
B	1.4
C	0.8
D	2.0
E	1.3
Total	6.0

One can change the amount of any ingredient without affecting the others. The only value affected is the total amount of the mixture, which may either increase or decrease depending on whether one is adding or subtracting an amount.

The experimental designs commonly used for nonmixture experiments are the Box-Wilson popularly known as the central composite, Plackett-Burman, and the Box-Behnken designs.

Mixture Experiments

This type pertains to experiments wherein the levels of each variable are restricted and dependent of each other. That is, the amount of each variable in the mixture should total 100%. Thus, changing the in-going levels of a variable in the blend will affect the levels of the other variables. By definition, Cornell (1981) stated that in the general mixture problem, the response that is measured is a function only of the proportions of the ingredients present in the mixture and is not a function of the amount of the mixture. However, there are experimental situations where the response depends on both the proportion and the amount. The solution to these situations are complex and the statistical formulation of the solution is discussed by Piepel and Cornell (1987).

Examples of mixture experiments include the blending of gasoline, cake mixes, beverages, aerosols, detergents, and cosmetic products. An example that is familiar to many is in textile manufacture:

Product	Ingredient	Percent
Shirt	Polyester	65
	Cotton	35

Note that if one increases polyester to 70%, one must reduce the cotton to 30%. Experimental designs useful in mixture experiments are the simplex and the McLean-Anderson (extreme vertices) designs.

Space Configuration of Nonmixture and Mixture Designs

There are differences in the response surface (mixture space) between the nonmixture and mixture designs. Fig. 8.3-1 shows the surface for two variables, which is a plane. Note that in the mixture design, only half of the square is the mixture space since $X_1 + X_2 = 1.0$. The response surface for three variables is given in Fig. 8.3-2. The triangle inside the box constitutes the surface of the mixture design. Again, note that $X_1 + X_2 + X_3 = 1.0$. Because of the constraint that the levels of all the variables in the blend must equal 1.0, the regression equation for a func-

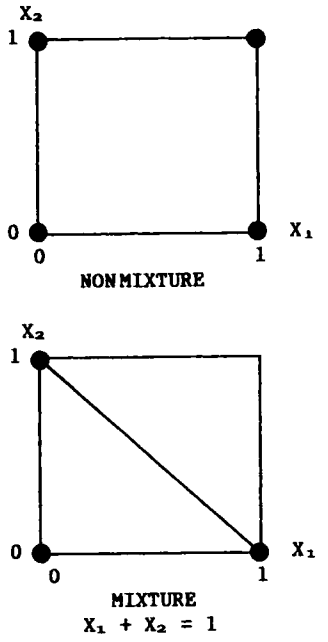


Fig. 8.3-1
Space configuration of nonmixture and mixture designs for two variables.

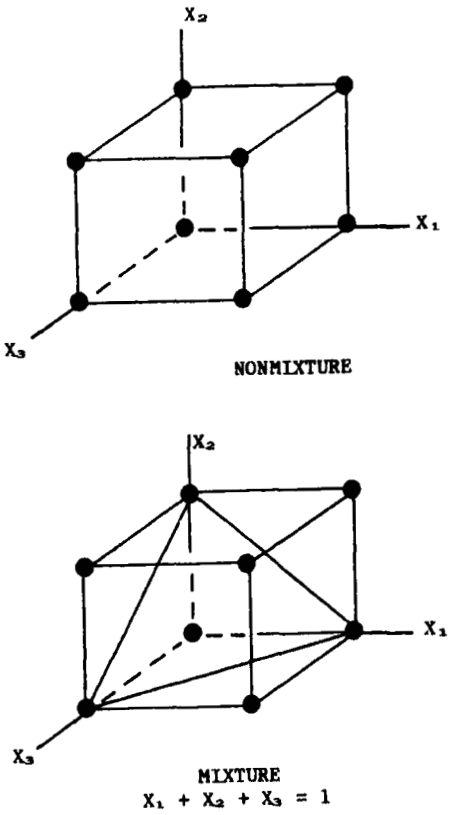


Fig. 8.3-2
Space configuration of nonmixture and mixture designs for three variables. Mixture space is the triangle shown at the bottom of the figure with vertices X_1, X_2, and X_3.

tion relating the dependent and the independent variables has no intercept. This constraint also results in regression coefficients interpreted differently.

Consider a 3-ingredient mixture model given by

$$Y_{ijk} = B_iX_i + B_{ij}X_iX_j + B_{ijk}X_iX_jX_k$$

The interpretation of the terms in the right hand side of the equation is summarized in Table 8.3-1. It is clear that the size of the regression coefficient is weighted by the coordinate of the design point. The contribution of each term to the response Y_{ijk} decreases with an increased number of ingredient combinations; the regression coefficient in the first term is the maximum effect itself; for the second term the

Table 8.3-1
Interpretation of the regression coefficients in mixture design.

Term	Maximum value of the variable	Relation between coefficient and maximum effect
B_iX_j	$X_i = 1$	$B_iX_j = 1B_i$
$B_{ij}X_iX_j$	$X_iX_j = (1/2)(1/2) = 1/4$	$B_{ij}X_iX_j = (1/4)B_{ui}$
$B_{ijk}X_iX_jX_k$	$X_iX_jX_k = (1/3)(1/3)(1/3) = 1/27$	$B_{ijk}X_iX_jX_k = (1/27)B_{ijk}$

effect is 1/4 of the coefficient, and for the third term it is 1/27. To illustrate, let us examine the plot of a two-ingredient design with response Y shown in Fig. 8.3-3. The amount of the curvature is given by $D = (1/4)b_{12}$.

Note that in mixture experiments, the B_{ij}s and higher orders are not called interaction in the usual regression terminology because the levels of the independent variables are not independent of each other. Positive value of B_{ij} indicates synergism of the effect of X_i and X_j, whereas a negative value indicates antagonism between them. Also, quadratic and cubic regression coefficients are to be interpreted as nonlinear blending effects. Figures 8.3-4–8.3-7 illustrate how to read a configuration of a mixture space. The figure is self-explanatory.

How does one evaluate the fit of the mixture model to the observed data? Due to the constraint that the proportion of the ingredients in the mixture adds up to 1.0, the mixture model has no intercept and as a result the calculated R-square criterion of fit is inflated, which often leads to misleading interpretation (Marquardt and Snee 1974; Kvalseth 1985). In fact, SAS output from PROC REG provides a warning message for the no-intercept model to redefine the R-square (Freund and Littell 1986).

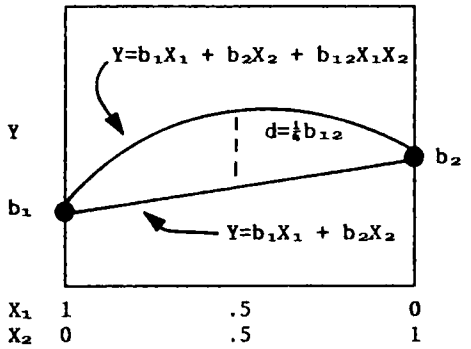


Fig. 8.3-3
An illustration of quadratic effect in mixture study with two variables.

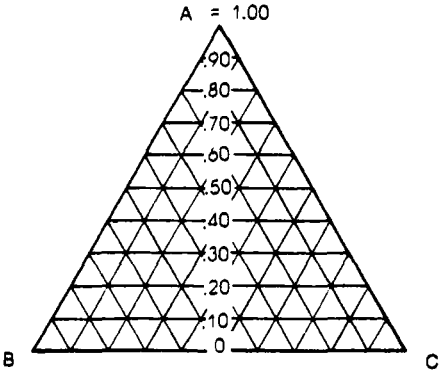


Fig. 8.3-4
Values for variable A in the mixture space.

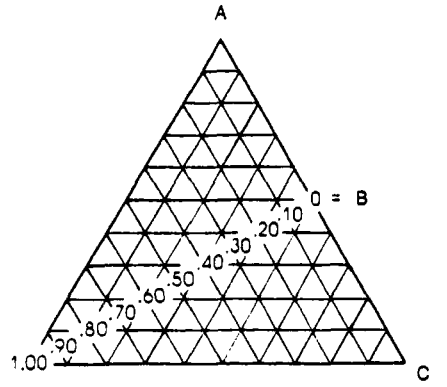


Fig. 8.3-5
Values for variable B in the mixture space.

There are other criteria that can be used to supplement the R-square statistic, such as the error or residual mean square in the analysis of variance and graphical analysis. All these criteria can be used to supplement each other. The examples that will follow illustrate the fitting of a model to the observed data.

Example 8.3-1

A two-variable mixture experiment was conducted involving two soap bases denoted by X_1 and X_2. Note that mixture study with two variables is known as binary blend. The proportion and the response Y of each blend are as follows:

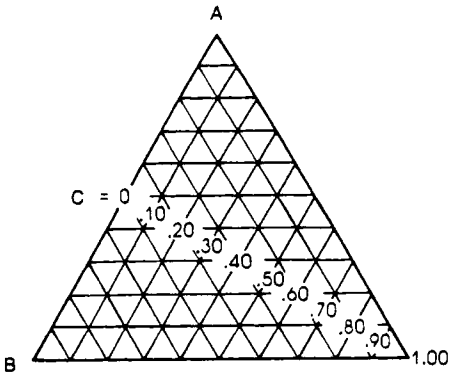


Fig. 8.3-6
Values for variable C in the mixture space.

X_1	X_2	Y
1.0	0.0	21.19
.9	.1	20.47
.8	.2	20.48
.7	.3	19.92
.6	.4	19.14
.5	.5	18.25
.4	.6	17.94
.3	.7	16.97
.2	.8	16.22
.1	.9	15.47
0.0	1.0	15.35

Notice the constraint $X_1 + X_2 = 1.0$. The response Y is the percentage of a certain fatty acid obtained from gas chromatograph analysis. It is hypothesized that the two bases will blend linearly so that the mixture model is

$$Y = B_1X_1 + B_2X_2 + \text{Error}.$$

However, before this model was used the plot of Y and the blend was done to verify the hypothesis (Fig. 8.3-8). The linear model appears appropriate as seen by the straight line plot of the data.

Table 8.3-2 shows the statistical analysis of the data for this example using the STATISTIX software. The fitted model is found to be

$$Y = 21.45X_1 + 15.17X_2$$

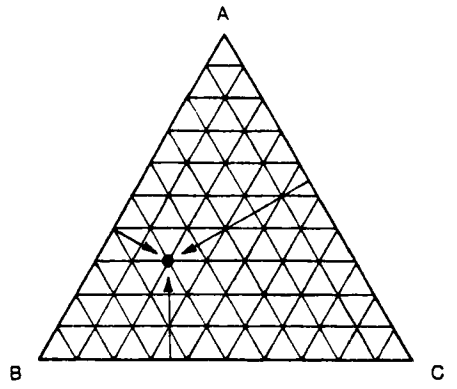


Fig. 8.3-7
An example of a point in the mixture space. This point has a coordinate of (A = 0.3, B = 0.5, C = 0.2).

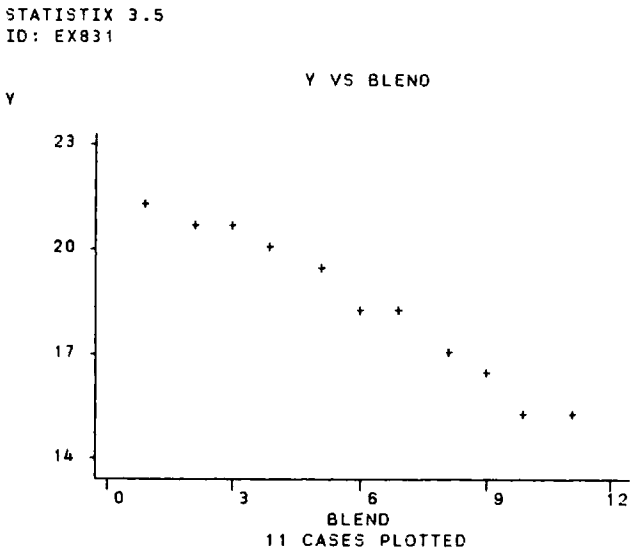


Fig. 8.3-8
Plot of response Y vs blend number for Example 8.3-1.

with adjusted R-square of 0.9998. Based on the size of the regression coefficients, variable X_1 is found to have greater effects on the response Y than variable X_2. Table 8.3-3 shows a summary output containing the raw data, fitted value of Y,

Table 8.3-2
Regression analysis of the soap base mixture experiment.

STATISTIX 3.5
ID: EX831

UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF Y

NOTE: MODEL FORCED THROUGH ORIGIN

PREDICTOR VARIABLES	COEFFICIENT	STD ERROR	STUDENT'S T	P
X1	21.449	1.5721E-01	136.44	0.0000
X2	15.169	1.5721E-01	96.49	0.0000

R SQUARED	0.9998	RESID. MEAN SQUARE (MSE)	7.767E-02
ADJUSTED R SQUARED	0.9998	STANDARD DEVIATION	2.787E-01

SOURCE	DF	SS	MS	F	P
REGRESSION	2	3730.8	1865.4	24016.49	0.0000
RESIDUAL	9	6.9905E-01	7.7672E-02		
TOTAL	10	3731.5			

CASES INCLUDED 11 MISSING CASES 0

Table 8.3-3
Estimates of fitted (predicted) and residual values for Example 8.3-1.

STATISTIX 3.5
ID: EX831

VIEW DATA

CASE	X1	X2	Y	FIT	RES
1	1.0000	0.0000	21.190	21.449	-0.2591
2	0.9000	0.1000	20.470	20.821	-0.3511
3	0.8000	0.2000	20.480	20.193	0.2869
4	0.7000	0.3000	19.920	19.565	0.3549
5	0.6000	0.4000	19.140	18.937	0.2029
6	0.5000	0.5000	18.250	18.309	-0.0591
7	0.4000	0.6000	17.940	17.681	0.2589
8	0.3000	0.7000	16.970	17.053	-0.0831
9	0.2000	0.8000	16.220	16.425	-0.2051
10	0.1000	0.9000	15.470	15.797	-0.3271
11	0.0000	1.0000	15.350	15.169	0.1809

and the residual. The plot of the residual versus Y indicates that the errors are evenly spread across blend suggesting that each observation is subject to the same random variation (Fig. 8.3-9).

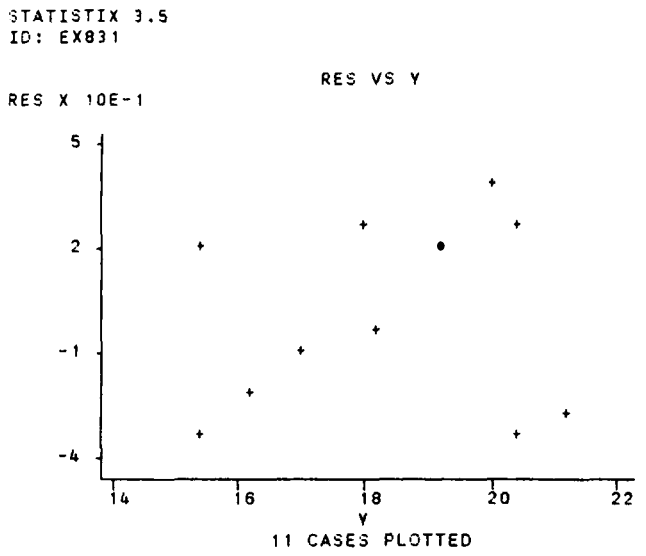


Fig. 8.3-9
Residual plot for Example 8.3-1.

Example 8.3-2

This is another example of a two-variable mixture problem to illustrate antagonism between variables. Hare (1974) reported the following data:

X_1	X_2	Y
1	0	14.7
0	1	35.5
2/3	1/3	17.5
1/3	2/3	24.0

In this example, X_1 is the proportion of stearine, X_2 the proportion of vegetable oil, and Y the solid fat index in 50 F. The STATISTIX output for these data is given in Table 8.3-4. The equation is found to be

$$Y = 14.77X_1 + 35.43X_2 - 19.59X_1X_2$$

Since the nonlinear blending coefficient, -19.59, is negative it indicates antagonism. The size of this effect is indicated by a vertical dash line in Fig. 8.3-10. If this effect was positive the nonlinear curve will be above the linear one and indicates synergism. The magnitude of antagonism when the proportion of each variable is equal to 0.5 is $(.5)(.5)(-19.59) = -4.90$, which is considerably less than the linear blending

Table 8.3-4
Regression analysis for mixture experiment to illustrate the nonlinear blending effects ($X_3 = X_1*X_2$).

```
STATISTIX 3.5
ID: EX832

VIEW DATA

 CASE        X1         X2         X3          Y

    1    1.0000     0.0000     0.0000     14.700
    2    0.0000     1.0000     0.0000     35.500
    3    0.6670     0.3330     0.2221     17.500
    4    0.3330     0.6670     0.2221     24.000


STATISTIX 3.5
ID: EX832

UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF Y

NOTE: MODEL FORCED THROUGH ORIGIN

PREDICTOR
VARIABLES     COEFFICIENT     STD ERROR      STUDENT'S T        P
---------     -----------     ---------      -----------     ------
X1              14.767        2.9231E-01       50.52        0.0126
X2              35.433        2.9231E-01      121.22        0.0053
X3             -19.585        1.3504          -14.50        0.0438

R SQUARED              1.0000     RESID. MEAN SQUARE (MSE)    8.996E-02
ADJUSTED R SQUARED     0.9998     STANDARD DEVIATION          2.999E-01

SOURCE        DF        SS            MS          F         P
------        --        --            --          -         -
REGRESSION     3      2358.5        786.17      8739.21    0.0079
RESIDUAL       1      8.9959E-02    8.9959E-02
TOTAL          3      2358.6

CASES INCLUDED 4    MISSING CASES 0
```

effects of either X_1 or X_2. Note that the size of the linear blending effects is the respective linear coefficient in the model.

Example 8.3-3

The data for this example is taken from the experiment used in Example 8.3-1, except that the response variable is melting temperature of the blend. The temperature Y corresponding to the 11 blends are as follows: 32.0, 23.9, 31.4, 29.6, 25.2, 26.7, 30.7, 27.7, 27.2, 31.7, 33.1.

The STATISTIX output is given in Table 8.3-5. The equation obtained is $Y = 27.845X_1 + 30.191X_2$ with an adjusted R-square of 0.9887 and a mean square er-

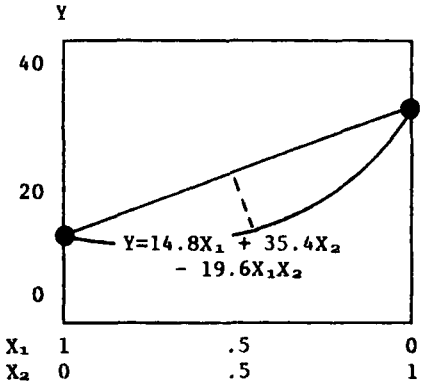


Fig. 8.3-10
Plot of response Y and blend to illustrate antagonism; X_1 = percent stearine, X_2 = percent vegetable oil, Y = percent solids at 50 F.

Table 8.3-5
Regression analysis for Example 8.3-3.

```
STATISTIX 3.5
ID: EX833

UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF Y

NOTE: MODEL FORCED THROUGH ORIGIN

PREDICTOR
VARIABLES     COEFFICIENT     STD ERROR      STUDENT'S T        P
---------     -----------     ---------      -----------     ------
X1              27.845         1.7510           15.90        0.0000
X2              30.191         1.7510           17.24        0.0000

R SQUARED                0.9907     RESID. MEAN SQUARE (MSE)    9.636
ADJUSTED R SQUARED       0.9887     STANDARD DEVIATION          3.104

SOURCE          DF        SS             MS          F         P
------          --      -------        -------     ------    ------
REGRESSION       2      9268.7         4634.3      480.93    0.0000
RESIDUAL         9      86.725         9.6361
TOTAL           10      9355.4

CASES INCLUDED 11   MISSING CASES 0
```

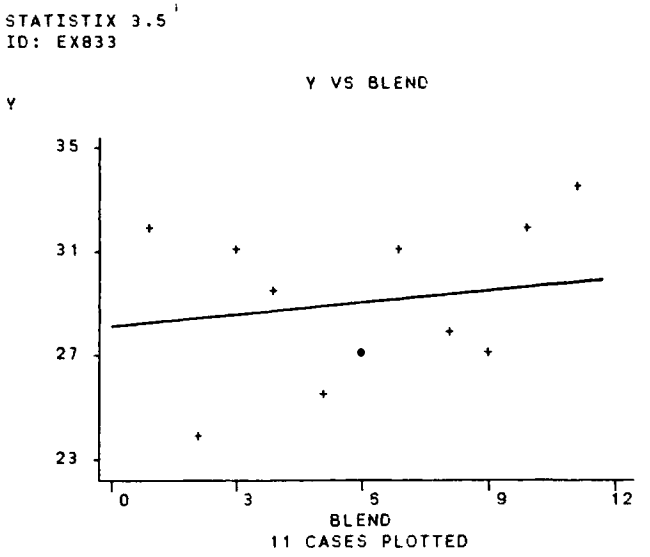


Fig. 8.3-11
Plot of melting temperature Y vs blend number for Example 8.3-3.

ror of 9.636. Although the R-square value is very high, the mean square error is also high, which indicates that the model is probably not a good fit. The plot shown in Fig. 8.3-11 reveals the divergent scatter of the data around the estimated regression line. This example demonstrates the inadequacy of the R-square statistic for assessing the no-intercept models. The inflated value of R-square is due to the use of uncorrected total sums of squares for the dependent variable Y. See Marquardt and Snee (1974), Cornell (1981), Draper and Smith (1981), Kvalseth (1985), and Freund and Littell (1986) for further reference.

8.4 PLACKETT AND BURMAN DESIGN

For simplicity, we shall cover only those designs where each factor is at low and high levels. At best, two levels are recommended for screening studies. The Plackett-Burman design (Plackett and Burman 1946) is an orthogonal fractional factorial where N, the total number of design points, is a multiple of 4; thus, we can have N equal to 4, 8, 12, 16, etc.

The Plackett-Burman design is suitably used to screen a large number of factors believed to be affecting important product characteristics or attributes, and is generally used during the initial phase of the study. Provided the interaction effects are nil or negligible, the Plackett-Burman design is effective for measuring main effects.

Being a fractional factorial design, the estimate of main effects will be contaminated with the interaction effects.

In this section, we consider designs with N = 8, 12, 16 for screening at least 7, 11, and 15 factors, respectively. Table 8.4-1 shows the initial block (design point) for obtaining the design matrix of the Plackett-Burman design. In this table, for simplicity the coefficient of 1 is not shown. Consider a design with N = 8 points. Using the initial block in Table 8.4-1, the remaining blocks of the design are found by cyclic enumeration. Writing down the initial block as

$$+ + + - + - -$$

then the second block is

$$- + + + - + -$$

and the third block, starting with the last sign of the second block, is

$$- - + + + - +$$

and so on. The result of the cyclic enumeration is given in Table 8.4-2. The total number of blocks is always N = k + 1 where the last block has a minus coefficient to provide an orthogonal design matrix. Note that if one is working with only 5 factors, columns F and G in Table 8.4-2 are deleted and the design would still be orthogonal.

Table 8.4-1
Initial block for cyclic construction of the Plackett-Burman design.

N	No. of factors $\leqslant$k	Initial block
8	7	+ + + − + − −
12	11	+ + − + + + − − − + −
16	15	+ + + + − + − + + − − + − − −
20	19	+ + − − + + + + − + − + − − − − + + −
24	23	+ + + + + − + − + + − − + + − − + − + − − − −

Source: Plackett and Burman (1946)

The estimate of the effect of each factor is found by a similar procedure discussed in Chapter 6 for calculating contrasts. That is, subtracting the sum of the responses with the plus sign from the sum of the responses with the minus sign, or in the form of a formula

Table 8.4-2
Design matrix for a 7-factor Plackett-Burman design.

Design point	Factor A	B	C	D	E	F	G		Response Y_i
1	+	+	+	−	+	−	−	Initial block	Y_1
2	−	+	+	+	−	+	−		Y_2
3	−	−	+	+	+	−	+	Cyclic construction	Y_3
4	+	−	−	+	+	+	−		Y_4
5	−	+	−	−	+	+	+		Y_5
6	+	−	+	−	−	+	+		Y_6
7	+	+	−	+	−	−	+		Y_7
8	−	−	−	−	−	−	−	Added block	Y_8

Note: Design points 2-7 obtained by cyclic construction.

$$\text{Average main effect} = [\Sigma Y(\text{plus sign}) - \Sigma Y(\text{minus sign})]/N \qquad (8.4\text{-}1)$$

Thus, the estimate of average main effect of factor A is

$$\text{Effect A} = [(Y_1 + Y_4 + Y_6 + Y_7) - (Y_2 + Y_3 + Y_5 + Y_8)]/8$$

and so on. As discussed in Section 8.1, the formula in matrix form for estimating factor effects is

$$b = (X'X)^{-1}X'Y \qquad (8.4\text{-}2)$$

where X is the design matrix, Y is the vector of responses, and b is the vector of parameter estimates.

The estimates of main effects can be positive or negative. Large value of effects indicates large influence of a factor in determining the value of the response. Factors with small effects may be left alone in product reformulation or entirely removed depending on the situation, i.e., their positive interaction with other ingredients in the formula. The statistical significance of effects can be obtained in the form of an analysis of variance shown in Table 8.4-3.

Test of significance for each regression coefficient is obtained by using the t-statistic. This is accomplished by first computing the common standard error SE from the MSE,

$$SE = \sqrt{MSE/N}$$

then $t = b/SE$ with $N-k-1$ DF. See Table D in the Appendix to determine the

Table 8.4-3
Analysis of variance table for the Plackett-Burman design.

Source of Variance	DF	SS	MS	F-ratio
Total	N−1	SSTO		
Regression	k	SSR	MSR	MSR/MSE
Residual	N−k−1	SSE	MSE	

Note:
SSTO = $\Sigma Y_i^2 - (\Sigma Y_i)^2/N$
SSR = b'X'Y, see also formula (8.1-6).
SSE = SSTO − SSR

significance of t. Note that when N = 8 and k = 7 the DF is zero. Thus it is suggested that in this case one may limit the study to 3 or 4 factors, unless the study is replicated.

Example 8.4-1

Five factors are suspected to affect the lather characteristics of a soap product. Due to constraints in equipment and cost, the researcher decides to eliminate two of the five factors for inclusion in the final product formulation. Table 8.4-4 shows the design matrix and the average lather score for each formulation (design point). The design matrix is taken from Table 8.4-2 for k = 5 factors with the coefficient of 1 indicated. In matrix form the solution is

Table 8.4-4
Data and calculations for Example 8.4-1.

Formulation	A	B	C	D	E	Y
1	+1	+1	+1	−1	+1	7.2
2	−1	+1	+1	+1	−1	6.5
3	−1	−1	+1	+1	+1	4.0
4	+1	−1	−1	+1	+1	3.8
5	−1	+1	−1	−1	+1	4.1
6	+1	−1	+1	−1	−1	3.4
7	+1	+1	−1	+1	−1	6.0
8	−1	−1	−1	−1	−1	3.0

$\Sigma Y_i = 38.0$ $\quad \bar{Y} = 4.750$ $\quad \Sigma Y_i^2 = 197.9$ $\quad i = 1, 2, .., 8$

Calculation of average main effects:
A = [(7.2 + 3.8 + 3.4 + 6.0) − (6.5 + 4.0 + 4.1 + 3.0)]/8 = 0.350
B = [(7.2 + 6.5 + 4.1 + 6.0) − (4.0 + 3.8 + 3.4 + 3.0)]/8 = 1.200
C = [(7.2 + 6.5 + 4.0 + 3.4) − (3.8 + 4.1 + 6.0 + 3.0)]/8 = 0.525
D = [(6.5 + 4.0 + 3.8 + 6.0) − (7.2 + 4.1 + 3.4 + 3.0)]/8 = 0.325
E = [(7.2 + 4.0 + 3.8 + 4.1) − (6.5 + 3.4 + 6.0 + 3.0)]/8 = 0.025

$$b = (X'X)^{-1})X'Y = \begin{bmatrix} b_0 \\ b_A \\ b_B \\ b_C \\ b_D \\ b_E \end{bmatrix} = \begin{bmatrix} 4.750 \\ 0.350 \\ 1.200 \\ 0.525 \\ 0.325 \\ 0.025 \end{bmatrix}$$

From the estimates of factor effects, it is concluded that the most important factor affecting lather is factor B, followed by C, A, D, and E. If factor E is believed to be technically nonfunctional, it can be removed from the formula.

The step-by-step procedure using X-STAT is shown in Table 8.4-5. The default of X-STAT is N = 16 design points; this means that the program can handle up to 15 factors. STATISTIX and SAS can also be used to analyze the data. Table 8.4-6 shows the design matrix generated by X-STAT for 11 and 14-factor studies.

8.5 BOX AND BEHNKEN DESIGN

In Chapter 4, the incomplete block design was presented and in Chapter 6, the 2^k factorial design was briefly discussed. The Box-Behnken designs (Box and Behnken 1960) are incomplete factorials formed by combining two-level factorial designs with incomplete block designs in a manner illustrated in Table 8.5-1. The resulting design is a rotatable second order used for estimating the coefficients of quadratic response surfaces. As shown in Table 8.5-1, the Box-Behnken design requires three levels of each factor to be studied. These levels are denoted by low (−1), medium (0), and high (1).

Table 8.5-2 shows the design matrix for some useful Box-Behnken designs. The row order (Run#) is printed at random, which is a desirable feature of the X-STAT software. This software provides analysis for three to seven factors. However, for large studies SAS is a better software to use. One disadvantage of this design is the large number of design points that results with large number of factors, thus its application to product formulation appears limited to three or four factors. For example, a 5-factor study requires 46 formulations that are prohibitive to conduct in many situations.

Table 8.4-5
Step-by-step analysis for Example 8.4-1 using X-STAT. The X-STAT commands are indicated by asterisk, i.e., * factor.

```
* factor
FACTOR[1]? x1
FACTOR[2]? x2
FACTOR[3]? x3
FACTOR[4]? x4
FACTOR[5]? x5
FACTOR[6]?
* design plackett-burman 8
8 experiment runs defined
*

* worksheet

                    Controlled Factors
RUN *--------*--------*--------*--------*--------*
 #      X1       X2       X3       X4       X5

 1     1.000    1.000    1.000   -1.000    1.000

 2     1.000    1.000   -1.000    1.000   -1.000

 3     1.000   -1.000    1.000   -1.000   -1.000

 4    -1.000    1.000   -1.000   -1.000    1.000

 5     1.000   -1.000   -1.000    1.000    1.000

 6    -1.000   -1.000    1.000    1.000    1.000

 7    -1.000    1.000    1.000    1.000   -1.000

 8    -1.000   -1.000   -1.000   -1.000   -1.000
*
 waiting

* char y
* y
Y[1]? 7.2
Y[2]? 6.0
Y[3]? 3.4
Y[4]? 4.1
Y[5]? 3.8
Y[6]? 4.0
Y[7]? 6.5
Y[8]? 3.0
```

Table 8.4-5 (*Continued*)

```
* fit y
Y:
  standard deviation about the regression = 0.9605
  explained variation about the mean (R-squared) = 89.40%
  condition of design matrix = 1.000
  model = LINEAR
* coefficient y

              Regression Coefficients for Y
              =============================

                                  Standard              Confidence
  Coefficient       Term            Error     T-Value   Coef <> 0

       4.750   1 (constant)        0.3396      13.99       94.8%
       0.3500  X1                  0.3396       1.031      58.4%
       1.200   X2                  0.3396       3.534      89.2%
       0.5250  X3                  0.3396       1.546      72.7%
       0.3250  X4                  0.3396      0.9571      55.6%
       0.02500 X5                  0.3396      0.0736      18.2%

Confidence figures are based on 2 degrees of freedom
*
 waiting

       1.200   X2                  0.3396       3.534      89.2%
       0.5250  X3                  0.3396       1.546      72.7%
       0.3250  X4                  0.3396      0.9571      55.6%
       0.02500 X5                  0.3396      0.0736      18.2%

Confidence figures are based on 2 degrees of freedom
* anova y

              Analysis of Variance for Y
              ==========================

  Source              df       SS        MS      F-Ratio

Total (corrected)      7     17.400
Regression             5     15.555    3.1110
                                                  3.372  (1)
Residual               2      1.845    0.9225

(1) Implies 74.6% confidence regression equation is nonzero.
*
 waiting
```

Table 8.4-6
Design matrix for 11 (A–K) and 14 (A–N) factors Plackett-Burman design.

Controlled Factors

RUN #	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	1.000	1.000	1.000	1.000	-1.000	1.000	-1.000	1.000	1.000	-1.000	-1.000	1.000	-1.000	-1.000
2	1.000	1.000	1.000	-1.000	1.000	-1.000	1.000	1.000	-1.000	-1.000	1.000	-1.000	-1.000	-1.000
3	1.000	1.000	-1.000	1.000	-1.000	1.000	1.000	-1.000	-1.000	1.000	-1.000	-1.000	-1.000	1.000
4	1.000	-1.000	1.000	-1.000	1.000	1.000	-1.000	-1.000	1.000	-1.000	-1.000	-1.000	1.000	1.000
5	-1.000	1.000	-1.000	1.000	1.000	-1.000	-1.000	1.000	-1.000	-1.000	-1.000	1.000	1.000	1.000
6	1.000	-1.000	1.000	1.000	-1.000	-1.000	1.000	-1.000	-1.000	-1.000	1.000	1.000	1.000	1.000
7	-1.000	1.000	1.000	-1.000	-1.000	1.000	-1.000	-1.000	-1.000	1.000	1.000	1.000	1.000	-1.000
8	1.000	1.000	-1.000	-1.000	1.000	-1.000	-1.000	-1.000	1.000	1.000	1.000	1.000	-1.000	1.000
9	1.000	-1.000	-1.000	1.000	-1.000	-1.000	-1.000	1.000	1.000	1.000	1.000	-1.000	1.000	-1.000
10	-1.000	-1.000	1.000	-1.000	-1.000	-1.000	1.000	1.000	1.000	1.000	-1.000	1.000	-1.000	1.000
11	-1.000	1.000	-1.000	-1.000	-1.000	1.000	1.000	1.000	1.000	-1.000	1.000	-1.000	1.000	1.000
12	1.000	-1.000	-1.000	-1.000	1.000	1.000	1.000	1.000	-1.000	1.000	-1.000	1.000	1.000	-1.000
13	-1.000	-1.000	-1.000	1.000	1.000	1.000	1.000	-1.000	1.000	-1.000	1.000	1.000	-1.000	-1.000
14	-1.000	-1.000	1.000	1.000	1.000	1.000	-1.000	1.000	-1.000	1.000	1.000	-1.000	-1.000	1.000
15	-1.000	1.000	1.000	1.000	1.000	-1.000	1.000	-1.000	1.000	1.000	-1.000	-1.000	1.000	-1.000
16	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000

Controlled Factors

RUN #	A	B	C	D	E	F	G	H	I	J	K
1	1.000	1.000	-1.000	1.000	1.000	1.000	-1.000	-1.000	-1.000	1.000	-1.000
2	1.000	-1.000	1.000	1.000	1.000	-1.000	-1.000	-1.000	1.000	-1.000	1.000
3	-1.000	1.000	1.000	1.000	-1.000	-1.000	-1.000	1.000	-1.000	1.000	1.000
4	1.000	1.000	1.000	-1.000	-1.000	-1.000	1.000	-1.000	1.000	1.000	-1.000
5	1.000	1.000	-1.000	-1.000	-1.000	1.000	-1.000	1.000	1.000	-1.000	1.000
6	1.000	-1.000	-1.000	-1.000	1.000	-1.000	1.000	1.000	-1.000	1.000	1.000
7	-1.000	-1.000	-1.000	1.000	-1.000	1.000	1.000	-1.000	1.000	1.000	1.000
8	-1.000	-1.000	1.000	-1.000	1.000	1.000	-1.000	1.000	1.000	1.000	-1.000
9	-1.000	1.000	-1.000	1.000	1.000	-1.000	1.000	1.000	1.000	-1.000	-1.000
10	1.000	-1.000	1.000	1.000	-1.000	1.000	1.000	1.000	-1.000	-1.000	-1.000
11	-1.000	1.000	1.000	-1.000	1.000	1.000	1.000	-1.000	-1.000	-1.000	1.000
12	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000

Table 8.5-1
Construction of the Box-Behnken design for k = 3 factors (A, B, C).

Incomplete block:				2^2 factorial:	
	A	B	C	X_i	X_j
Row					
1	x	x		−1	−1
2	x		x	1	−1
3		x	x	−1	1
				1	1

Replace row 1 by the factorial matrix and put zeros on others:

A	B	C
−1	−1	0
1	−1	0
−1	1	0
1	1	0

Replace row 2 similarly:

−1	0	−1
1	0	−1
−1	0	1
1	0	1

Replace row 3 similarly:

0	−1	−1
0	1	−1
0	−1	1
0	1	1

Center point:

0	0	0
0	0	0
0	0	0

Note: X_i = factor 1, ith level.
X_j = factor 2, jth level.

8.6 BOX AND WILSON DESIGN

Since its publication in 1951, the Box-Wilson design (Box and Wilson 1951) is the most widely used design in industrial applications for searching optimum operating conditions, i.e., optimal areas in contour maps. Some applications of the design include those of Jao *et al.* (1982) on meat loaf analog system, McLellan *et al.* (1984) on carbonated apple juice study, Diptee *et al.* (1989) on protein extraction studies from brewer's spent grain, and Galvez *et al.* (1990) on processing peanut beverage.

Table 8.5-2
Design matrix for 3 (A-C) and 4 (A-D) factors Box-Behnken design.

RUN #	Controlled Factors A	B	C
1	1.000	1.000	0.000
2	1.000	-1.000	0.000
3	-1.000	1.000	0.000
4	-1.000	-1.000	0.000
5	1.000	0.000	1.000
6	1.000	0.000	-1.000
7	-1.000	0.000	1.000
8	-1.000	0.000	-1.000
9	0.000	1.000	1.000
10	0.000	1.000	-1.000
11	0.000	-1.000	1.000
12	0.000	-1.000	-1.000
13	0.000	0.000	0.000
14	0.000	0.000	0.000
15	0.000	0.000	0.000

RUN #	Controlled Factors A	B	C	D
1	1.000	1.000	0.000	0.000
2	1.000	-1.000	0.000	0.000
3	-1.000	1.000	0.000	0.000
4	-1.000	-1.000	0.000	0.000
5	0.000	0.000	1.000	1.000
6	0.000	0.000	1.000	-1.000
7	0.000	0.000	-1.000	1.000
8	0.000	0.000	-1.000	-1.000
9	0.000	0.000	0.000	0.000
10	1.000	0.000	0.000	1.000
11	1.000	0.000	0.000	-1.000
12	-1.000	0.000	0.000	1.000
13	-1.000	0.000	0.000	-1.000
14	0.000	1.000	1.000	0.000
15	0.000	1.000	-1.000	0.000
16	0.000	-1.000	1.000	0.000
17	0.000	-1.000	-1.000	0.000
18	0.000	0.000	0.000	0.000
19	1.000	0.000	1.000	0.000
20	1.000	0.000	-1.000	0.000
21	-1.000	0.000	1.000	0.000
22	-1.000	0.000	-1.000	0.000
23	0.000	1.000	0.000	1.000
24	0.000	1.000	0.000	-1.000
25	0.000	-1.000	0.000	1.000
26	0.000	-1.000	0.000	-1.000
27	0.000	0.000	0.000	0.000

Like the Box-Behnken design, it is primarily used for the exploration of quadratic response surfaces, the most popular being the central composite rotatable design. Briefly, this design consists of a 2-level factorial augmented by a replicated center point. Figure 8.6-1 shows the configuration of a three-variable composite design with M = 15 design points. It can be easily seen that the first eight points correspond to the coordinates of the 2^3 factorial. For example, a plane with vertices of points 1, 2, 3, and 4 form the four points of the 2^2 factorial with an additional point 9 in the center; the diagram is shown in Fig. 8.6-2. The length of the axial portion denoted by α may vary from 1.41 to 4.00 to satisfy orthogonality and/or rotatability (Box and Hunter 1957). The choice of the axial length for various number of factors is given in Table 8.6-1. The popular choice for orthogonality is $\alpha = 2$. Table 8.6-2 shows the design matrix for a three-factor composite design.

Example 8.6-1

The purpose of this study was to find the optimum levels of salt and dextrose that will result in high consumer acceptance by masking the effect of perceived saltiness in the product. Consumer acceptance was measured by liking for saltiness and overall flavor. As discussed before, saltiness and overall flavor are called response variables, whereas the amounts of salt and dextrose are called variables or factors.

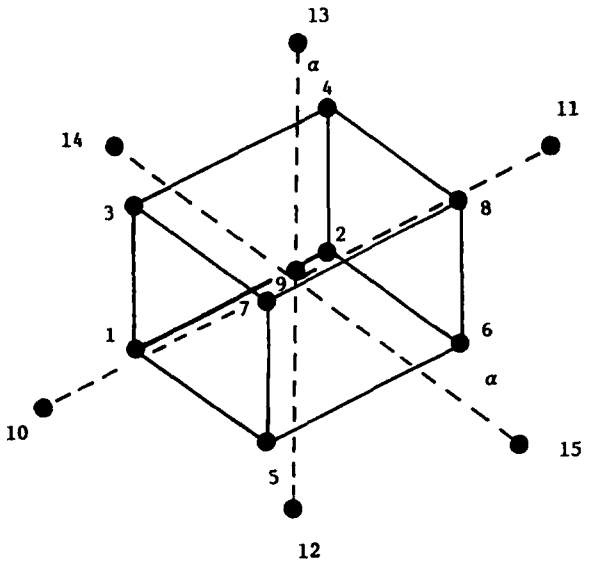


Fig. 8.6-1
The 15 design points of a three-variable composite design. The length of the broken dash line is α.

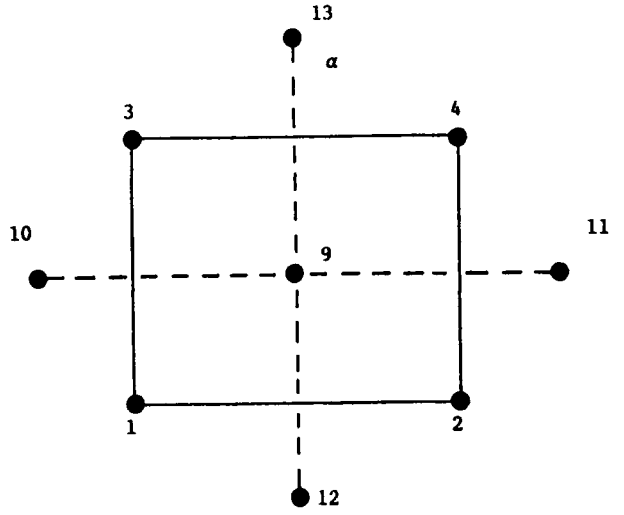


Fig. 8.6-2
Side of a three-factor composite design. The length of the broken dash line from the center is α.

Table 8.6-1
Composite rotatable designs satisfying orthogonality.

Number of factors k	Factorial points k_c	Axial points 2k	Center points k_0	Total number of observations M	α
2	4	4	8	16	1.414
3	8	6	9	23	1.682
4	16	8	12	36	2.000
5	32	10	17	59	2.378
5(1/2)	16	10	10	36	2.000
6(1/2)	32	12	15	59	2.378
7(1/2)	64	14	222	100	2.828

Note: $M = k_c + 2k + k_0$
$\alpha = k_c^{1/4}$
Number of factors for k = 5, 6, 7 are 1/2 fractional factorial discussed in Chapter 6.
Source: Box and Hunter (1957).

The lower and upper limits of the variables were determined using a model system consisting of distilled water as the medium where various amounts of salt and dextrose were added to it. It is important that the levels of these variables must be perceptually distinguishable. Based on the results from the model system, the limits were near to the following values:

Table 8.6-2
Design matrix (points) of a three-factor composite design.

Formulation	Ingredients X_1	X_2	X_3	
1	−1	−1	−1	
2	1	−1	−1	
3	−1	1	−1	
4	1	1	−1	
5	−1	−1	1	2^k portion
6	1	−1	1	
7	−1	1	1	
8	1	1	1	
9	0	0	0	Center point
10	−2	0	0	
11	2	0	0	
12	0	−2	0	2k portion
13	0	2	0	
14	0	0	−2	
15	0	0	2	

Note: k = number of factors (independent variables).
Formulations 1–8 form the 2^k portion.
Formulation 9 is the center point.
Formulations 10–15 form the 2^k portion.

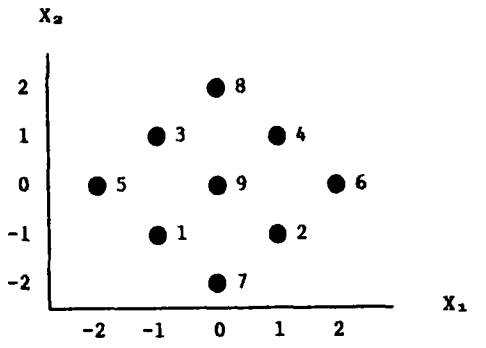
	Lower limit	Upper limit
% salt (X_1)	1.4	3.0
% dextrose (X_2)	1.4	10.2

A 2-factor central composite design was used in the study. With two independent variables, the design consists of nine experimental points as shown in Fig. 8.6-3. The relationship between the coded levels and the actual levels is given by (See also Eq. 8.1-4)

$$X_i = (L - C)/S, \qquad i = 1, 2$$

where L is the level of interest, C is the center level, and S is the so-called scaling factor. For example, when $X_1 = -1$

$$X_1 = (1.8 - 2.2)/0.4 = -1 \text{ for salt.}$$



X_1	1.4	1.8	2.2	2.6	3.0	S = 0.4
CODED LEVEL	-2	-1	0	1	2	
X_2	1.4	3.6	5.8	8.0	10.2	S = 2.2

Fig. 8.6-3
Coded and actual levels for two-factor composite design. S = scaling factor.

Likewise

$$X_2 = (3.6 - 5.8)/2.2 = -1 \text{ for dextrose.}$$

Based on the coded levels, the design matrix is shown in Table 8.6-3 as well as the matrix corresponding to the actual levels. The coded or the actual levels are the input data in most computer programs, thus knowing these levels facilitates computer analysis. The actual levels provide the experimenter with the level combinations necessary to make the product. Note that in this table, $\alpha = \pm 2$. Depending on the design, other values of α may be used (See Table 8.6-1).

The next step is the sensory evaluation of the nine formulations plus the control sample. It is very important in optimization study to include a control sample as a base point for comparison. A total of 108 panelists were used to evaluate 10 formulations (9 test formulations plus the control). To reduce the number of samples to be evaluated by each panelist, a balanced incomplete block design augmented with a control in every block was used as described in Chapter 4 and in Gacula and Singh (1984). Each panelist evaluated two test formulations plus the control using the 9-point hedonic scale. The results are shown in Table 8.6-4.

Using the data in this table, a quadratic model of the form

$$Y = B_0 + B_1X_1 + B_2X_2 + B_{11}X_1^2 + B_{22}X_2^2 + B_{12}X_{12} + \text{error}$$

Table 8.6-3
Design matrix for a 2-factor composite design for Example 8.6-1.

Formulation	X_1	X_2	X_1	X_2
	Coded levels		Actual levels	
1	−1	−1	1.8	3.6
2	1	−1	2.6	3.6
3	−1	1	1.8	8.0
4	1	1	2.6	8.0
5	−2	0	1.4	5.8
6	2	0	3.0	5.8
7	0	−2	2.2	1.4
8	0	2	2.2	10.2
9	0	0	2.2	5.8

Table 8.6-4
Adjusted mean scores for nine formulations and the control.

Formulation	Saltiness	Overall flavor
1	5.70	5.91
2	6.31	6.40
3	4.90	5.15
4	5.68	5.02
5	4.86	4.80
6	5.54	5.74
7	5.75	5.63
8	4.88	4.94
9	5.63	5.70
Control	6.00	5.90
Standard error	0.252	0.288

Note: Adjusted by intrablock analysis, see Chapter 4.

was used. This model includes the linear effects B_1 and B_2, the quadratic effects B_{11} and B_{22}, the interaction effect B_{12}, and the random error component. The resultant equations for saltiness and for overall flavor are as follows:

$$\text{Saltiness } Y_1 = 5.81 + 0.229X_1 - 0.264X_2 - 0.141X_1^2 - 0.113X_2^2 + 0.043X_1X_2$$

$$\text{Overall flavor } Y_2 = 5.81 + 0.187X_1 - 0.293X_2 - 0.129X_1^2 - 0.126X_2^2 - 0.155X_1X_2$$

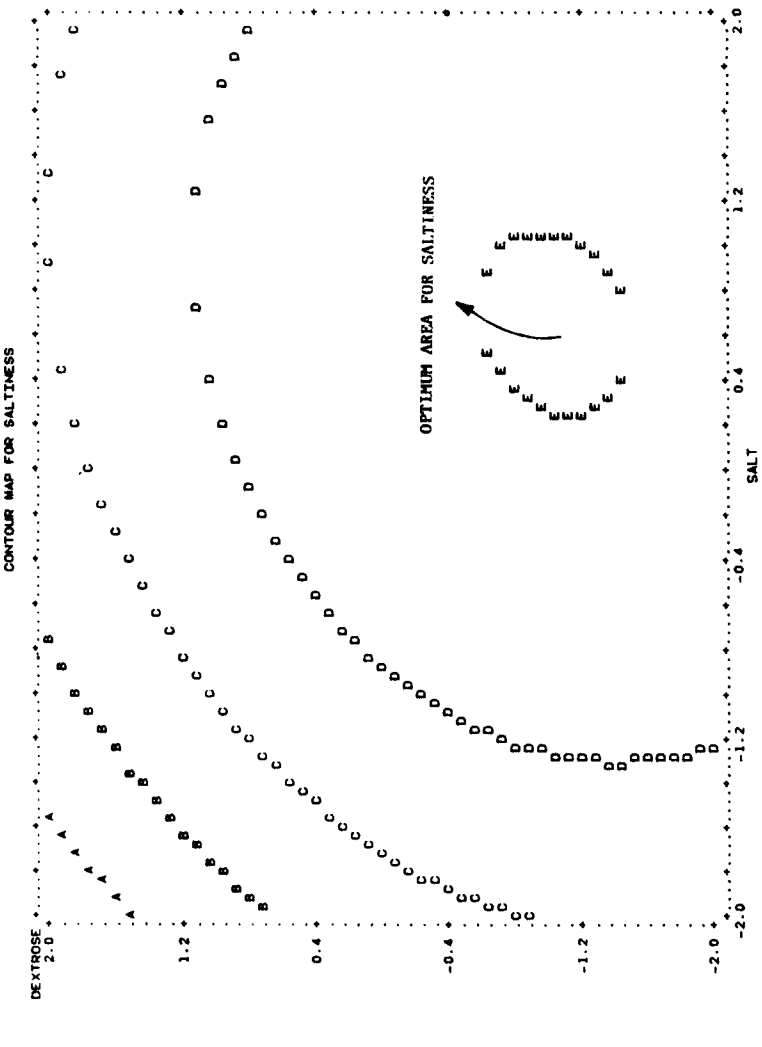


Fig. 8.6-4
Contour map for saltiness obtained from a quadratic model. Each contour line denoted by A, B, C . . , represents a value of the response variable.

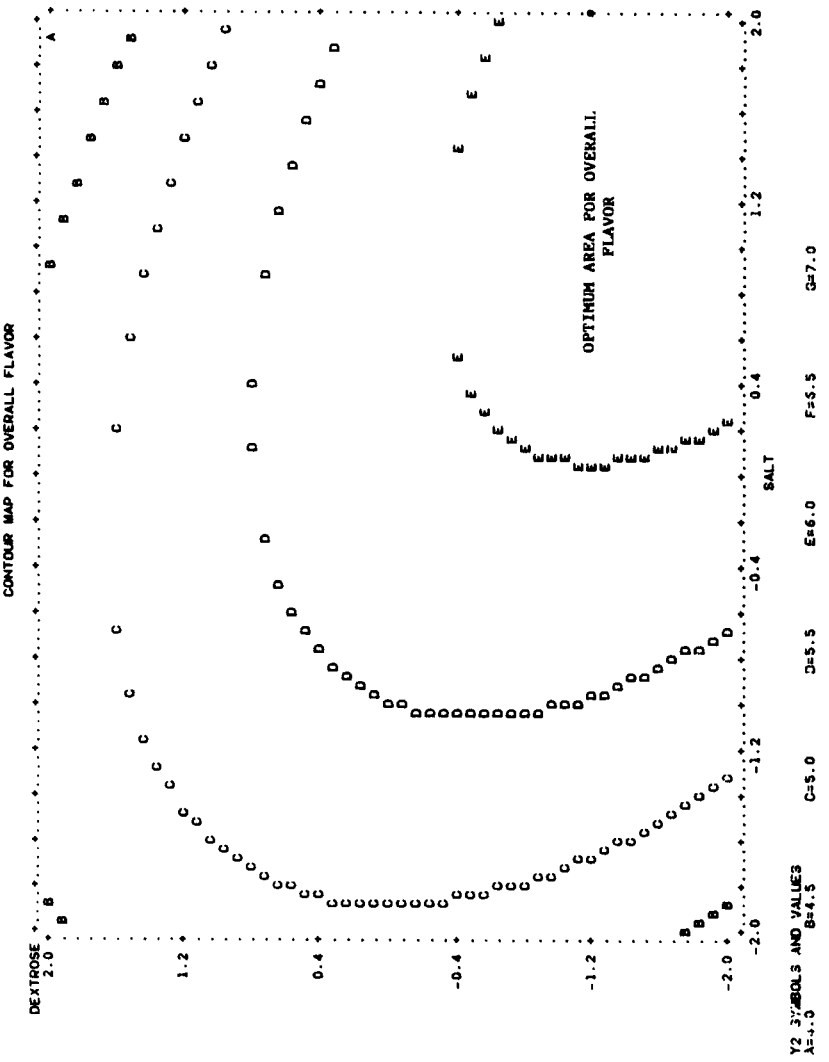


Fig. 8.6-5
Contour map for overall flavor.

The contour maps based on the above equations are shown in Figures 8.6-4 and 8.6-5 for saltiness and overall flavor, respectively. Both contour maps depict a "mountain." A contour line may be selected on the map to obtain desirable levels of salt and dextrose. For example, for saltiness with a predicted score of around 6, the optimum levels are 2.5% salt and 1.8% dextrose. For overall flavor, the levels are outside the experimental region, but a conservative value would be 3.0% salt and 1.4% dextrose.

Considering the direction of panelist responses, the optimum area is located on the lower right corner of the map; thus two formulas were selected on this map for consumer test:

	Formula 1	Formula 2
% salt	2.7	3.0
% dextrose	3.2	1.4

The current formula (Control) contains 1.8% salt and 1.5% dextrose. Note that higher level of salt is necessary for microbial stability of the product.

The data for saltiness and overall flavor can be linearly combined and standardized by dividing each term of the linear combination by their respective standard errors. In doing so, one contour map is obtained for both responses facilitating a simple interpretation of the map. However, there are constraints in combining responses. In our example, we are fortunate that both responses have similar contour maps; had the map been not similar, it would be difficult to combine them, unless a compromise is made between the two responses. The formula for combining responses is

$$Y_i = \Sigma[(X_{1i} - M_1)/S_1 + (X_{2i} - M_2)/S_2]$$

where Y_i = standardized combined response, X_{1i} = observation in the ith design point (i = 1, 2, .., 9) for variable X_1, M_1 = grand mean for variable X_1. and S_1 = standard error for variable X_1; similar definition applies to variable X_2. Assuming the same scale is used, response variable with small standard error relative to the other variables would contribute more to the value of Y_i. If the standard errors among the variables are close to each other, the effect of each variable will be weighted equally.

In our example, the standard error for saltiness is found to be 0.252 and for flavor, 0.288. Since they are close, for practical purposes both variables contribute equally to the value of Y_i. Table 8.6-5 shows the SAS program used for the analysis of the combined response, and the resulting output shown in Table 8.6-6. The SAS program to generate the map in Fig. 8.6-6 is given in Table 8.6-7.

As shown in Fig. 8.6-6 the results for the combined analysis indicate that Formula 1 falls nicely within the optimum area; the % salt for Formula 2 is on the

Table 8.6-5
SAS program (PROC RSREG) for obtaining an optimization regression equation.

```
DATA;
INPUT X1-X2 Y;
CARDS;
-1   -1    2.296
 1   -1    6.529
-1    1   -3.431
 1    1   -0.744
-2    0   -4.782
 2    0    1.183
 0   -2    1.652
 0    2   -4.224
 0    0    1.411
PROC RSREG;
MODEL Y = X1-X2;
TITLE 'TOTAL STANDARDIZED RESPONSE';
```

Note: PROC RSREG automatically provides a full model, i.e., main effects, interaction effects, and quadratic effects are estimated. Solution of optimum combination of factors is automatically computed.

borderline of the area. Since both saltiness and overall flavor were weighted equally as a result of their similar standard errors, the resulting total standardized contour map in Fig. 8.6-6 would optimize simultaneously the two response variables in one map.

The final phase of the optimization study is a consumer test to verify the performance of the selected formulas against the control. In this study, 110 panelists participated in the test. Based on the 9-point hedonic scale, the mean scores obtained in the consumer test are as follows:

	Control	Formula 1	Formula 2
Overall flavor	6.75	7.09	7.09
Saltiness	6.80	7.05	6.70
Texture	6.82	7.05	7.26*
Juiciness	6.44	6.95	7.11*

Notice that Formula 2 is significantly (*$p = 0.05$) more well-liked in texture and juiciness than the control. Based on the mean scores, both Formulas 1 and 2 are satisfactory. If manufacturing constraints and other considerations would prevent the production of either formula, then the contour map should be consulted noting that the formula to be selected is within the optimum area.

Table 8.6-6
SAS output for PROC RSREG. Note solution for optimum response at the bottom of the table.

TOTAL STANDARDIZED RESPONSE ANALYSIS

RESPONSE SURFACE FOR VARIABLE Y

RESPONSE MEAN	-0.0122222
ROOT MSE	1.940195
R-SQUARE	0.8945092
COEF OF VARIATION	999.99

REGRESSION	DF	TYPE I SS	R-SQUARE	F-RATIO	PROB
LINEAR	2	80.66533367	0.7535	10.71	0.0430
QUADRATIC	2	14.49669603	0.1354	1.93	0.2898
CROSSPRODUCT	1	0.59752900	0.0056	0.16	0.7170
TOTAL REGRESS	5	95.75955869	0.8945	5.09	0.1055

RESIDUAL	DF	SS	MEAN SQUARE
TOTAL ERROR	3	11.29306486	3.76435495

PARAMETER	DF	ESTIMATE	STD DEV	T-RATIO	PROB
INTERCEPT	1	2.50288889	1.44613565	1.73	0.1819
X1	1	1.57083333	0.56008593	2.80	0.0676
X2	1	-2.06266667	0.56008593	-3.68	0.0347
X1*X1	1	-1.00735417	0.54230087	-1.86	0.1602
X2*X1	1	-0.38650000	0.97009728	-0.40	0.7170
X2*X2	1	-0.87897917	0.54230087	-1.62	0.2035

FACTOR	DF	SS	MEAN SQUARE	F-RATIO	PROB
X1	3	43.1967	14.3989	3.83	0.1498
X2	3	61.54199	20.514	5.45	0.0987

SOLUTION FOR OPTIMUM RESPONSE

FACTOR	CRITICAL VALUE
X1	1.04901828
X2	-1.40396515

PREDICTED VALUE AT OPTIMUM 4.774761

EIGENVALUES	EIGENVECTORS X1	X2
-0.739536	-0.585143	0.8109299
-1.1468	0.8109299	0.5851433

SOLUTION WAS A MAXIMUM

Table 8.6-7
SAS program (PROC RSP) for obtaining contour map.

```
PROC RSP;
MODEL Y = 2.502 + 1.572*X1 - 2.063*X2 - 1.01*X1*X1
          - .387*X1*X2 - .879*X2*X2;
RANGES X1 = -2 TO 2 BY .5
       X2 = -2 TO 2 BY .5
        Y = -5 TO 7 BY 1;
PLOT X1*X2;
TITLE 'TOTAL STANDARDIZED RESPONSE';
```

8.7 MIXTURE DESIGNS

The nature of mixture experimentation was introduced in Section 8.3. Many industrial products such as detergents, cake mixes, cosmetics, and beverages are made by mixing a number of ingredients into a blend. The number of ingredients in a given product can go as high as 10 components in one mixture. Recall that in mixture experiments, the effect of each ingredient is assumed to depend only on their relative proportions in the blend or mixture as discussed in Section 8.3. In this section, we discuss several mixture designs useful in product formulation.

Mixture designs are based on a simplex configuration first used by Claringbold (1955) in a study of the joint action of related hormones in animal experimentation. In mathematics, a simplex is an n-dimensional line with vertices and faces. For example, a simplex of dimension 1 is a straight line with 2 vertices; a simplex of dimension 2 is a triangle with 3 vertices; and a simplex of dimension 3 is a tetrahedron. Figure 8.7-1 shows different types of simplices. Note that the side of a simplex is called a face.

Like the Box-Wilson and the Box-Behnken designs, the use of the mixture designs is to obtain optimum formulation of mixtures of ingredients to meet certain objectives. As such, the interpretation of the results of mixture experiments is in many ways similar to the response surface method. However, since the sum of the proportions for all components in the mixture must equal 1.0 or 100%, the response surface designs no longer apply. In this section, we first consider a simple mixture problem where all ingredients in the blend are not constrained. They can take on values from 0 to 1.0 or 0 to 100%, i.e.,

$$0 \leqslant X_i \leqslant 1, \qquad i = 1,2, .., q \tag{8.7-1}$$

where X_i is the proportion of the ith ingredient in the blend and $X_1 + X_2 + .. + X_q = 1$.

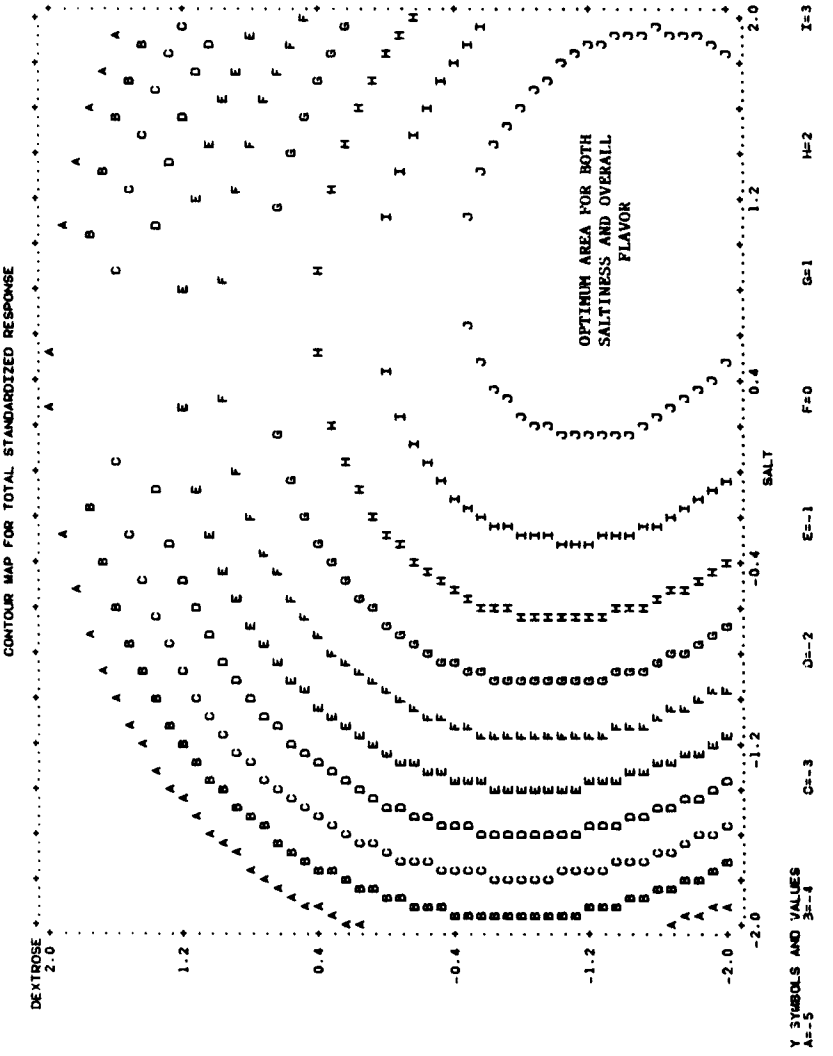
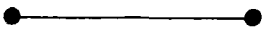
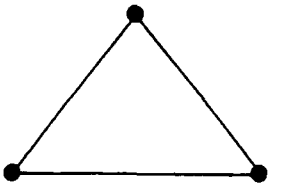


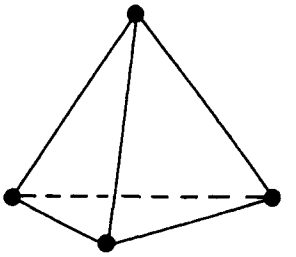
Fig. 8.6-6
Contour map for the combined responses weighted by their respective standard errors.



SIMPLEX OF DIMENSION 1



SIMPLEX OF DIMENSION 2



SIMPLEX OF DIMENSION 3

Fig. 8.7-1
Configuration of simplices

In most product formulation work, such as in the personal care, household, and the food industries, a formulation may consist of 8 or more ingredients. Often one may desire to work only with 3 of the 8 ingredients. The constraint that $\Sigma X_i = 1.0$ can still be achieved by an appropriate scaling so that $X_1 + X_2 + X_3 = 1.0$. The scaling is possible since we are dealing with proportions.

For example, suppose we have a formulation consisting of only 4 ingredients — X_1, X_2, X_3, and X_4, and the total is obviously equal to 1.0. Suppose that the first two ingredients are to be studied and that they comprise 30% of the total. A simple mixture design shown below can be used:

Blend	X_1	X_2
1	1.0	0.0
2	0.0	1.0
3	0.67	0.33
4	0.33	0.67

In this example, the remaining ingredients provide the 70% of the total mixture divided according to the following proportions: $X_3 = 50\%$ and $X_4 = 20\%$. Note that the mixture design satisfies the constraint that $X_1 + X_2 = 1.0$. Since we are working only with 30% of the total mixture, the blend to be formulated should be scaled accordingly to reflect the appropriate proportions by multiplying the design coordinate above by 0.30, resulting in the following formulations:

Blend	X_1	X_2	X_3	X_4
1	0.30	0.0	0.50	0.20
2	0.0	0.30	0.50	0.20
3	0.20	0.10	0.50	0.20
4	0.10	0.20	0.50	0.20

Note that $X_1 + X_2 + X_3 + X_4 = 1.0$. In producing the product blend, blend 1 would consist of 30% X_1, 0% X_2, 50% X_3, 20% X_4, and so on.

Mixture Models

As a result of the constraint that $\Sigma X_i = 1.0$, mixture model has no constant term B_o (intercept) and squared terms $B_{ij}X_i^2$. As shown by Scheffé (1958), data from the mixture experiment can be fitted by a polynomial regression in one of the following models:

Linear:

$$E(Y) = \Sigma B_i X_i$$

Quadratic:

$$E(Y) = \Sigma B_i X_i + \Sigma B_{ij} X_i X_j$$

Special cubic:

$$E(Y) = \Sigma B_i X_i + \Sigma B_{ij} X_i X_j + \Sigma B_{ijk} X_i X_j X_k$$

Full cubic:

$$E(Y) = \Sigma B_i X_i + \Sigma B_{ij} X_i X_j + \Sigma B_{ij} X_i X_j (X_i - X_j) + \Sigma B_{ijk} X_i X_j X_k$$

The notation, E(Y), indicates expected value of the response. The interpretation of the terms on the right hand side of the model equation is as follows:

$B_i X_i$ = response due to the pure components and B_i is the linear coefficient.

$B_{ij}X_iX_j$ = response due to synergism of the binary mixtures and B_{ij} is the quadratic coefficient.

$B_{ij}X_iX_j(X_i - X_j)$ = response due to synergism of the binary mixtures and B_{ij} is the cubic coefficient.

$B_{ijk}X_iX_jX_k$ = response due to synergism of the ternary mixtures for components i, j and k; B_{ijk} is the cubic coefficient.

Note that the pure components and the binary mixtures are needed to estimate the coefficients of the quadratic model. Suppose q = 3 ingredients. The mixture models can be one of the following:

Linear:

$$E(Y) = B_1X_1 + B_2X_2 + B_3X_3$$

Quadratic:

$$E(Y) = \text{Linear} + B_{12}X_1X_2 + B_{13}X_1X_3 + B_{23}X_2X_3$$

Special cubic:

$$E(Y) = \text{Linear} + \text{quadratic} + B_{123}X_1X_2X_3$$

Full cubic:

$$E(Y) = \text{Linear} + \text{quadratic} + \text{special cubic} + b_{12}X_1X_2(X_1 - X_2) + B_{13}X_1X_3(X_1 - X_3) + B_{23}X_2X_3(X_2 - X_3)$$

In practice, the linear and the quadratic models are the most useful. The estimation of the various coefficients of the model can be done using a calculator when the observations are taken within the simplex. See numerical examples in Cornell (1983) and Snee (1971). However, the wide availability of multiple regression programs favors computer analysis.

Scheffé Simplex-Lattice Design

The design coordinate for each ingredient in the simplex-lattice design (Scheffé 1958) is determined by the formula

$$X_i = 0, 1/m, 2/m, .., 1$$

where i = 1, 2, .., q ingredients. The proportion of each ingredient takes m + 1 equally spaced values from 0 to 1. A good property of this design is that the equally-spaced distribution of points over the simplex provides a desirable form and enough points to conduct polynomial regression analysis.

For example, if m = 3, the possible coordinates of a blend for X_1, X_2, and X_3 are, respectively, 0, 1/3, and 2/3. The total number of design points (Gorman and Hinman 1962) is given by

$$\begin{aligned} M &= (m + q - 1)!/m!(q - 1)! \\ &= q(q+1) \,..\, (q+m-1)/(1)(2) \,..\, (m) \end{aligned}$$

where the symbol ! denotes a "factorial operation," i.e., m! = m(m−1)(m−2) .. (m). Thus if we have three ingredients (q = 3) the total number of blends in the simplex-lattice design is

$$M = 3(4)(5)/(1)(2)(3) = 10$$

Since the design can have several combinations of q and m, they can be distinguished from each other by the name (q,m) lattice design. In our example, this is a (3,3) simplex-lattice design and its configuration is given in Fig. 8.7-2c.

Figure 8.7-2a shows the diagram of a (3,2) simplex-lattice with six design points equally spaced on the faces of the simplex at $X_i = 0, 1/2$, and 1.0. As shown in this figure, design point 1 has a coordinate of $(X_1, X_2, X_3) = (1, 0, 0)$ and likewise design point 6 has $(X_1, X_2, X_3) = (1/2, 0, 1/2)$. These points are appropriately collected to form the design coordinates shown in Table 8.7-1. As discussed in Section 8.1, these coordinates are used in the least squares estimation of the parameters of a postulated model. Together with the measured responses, the design coordinates are input data in computer analysis. It is again emphasized to readers the important role of design coordinates in statistical analysis of mixture data.

One model for fitting data from a 3-ingredient simplex-lattice design is the special cubic

$$Y = B_1X_1 + B_2X_2 + B_3X_3 + B_{12}X_1X_2 + B_{13}X_1X_3 + B_{23}X_2X_3 + B_{123}X_1X_2X_3 + \text{Random error}$$

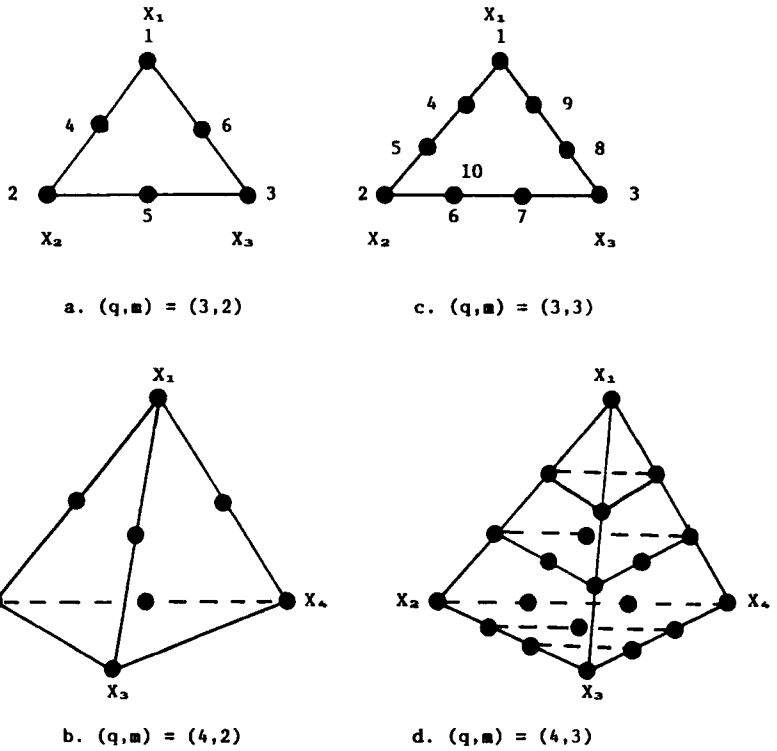


Fig. 8.7-2
Space configuration of simplex lattice designs.

Table 8.7-1
Design coordinates for simplex-lattice involving three ingredients.

Design Point	X_1	X_2	X_3
1	1	0	0
2	0	1	0
3	0	0	1
4	1/2	1/2	0
5	0	1/2	1/2
6	1/2	0	1/2

Note: Since there are only 6 design points, only the pure component effects and the binary mixture effects can be estimated.

where, as discussed earlier, B_1, B_2, and B_3 are regression coefficients for the pure components; B_{12}, B_{13}, and B_{23} are regression coefficients for the quadratic components, and B_{123} the cubic coefficient. Note that there are seven unknown parameters to be estimated, and one must choose a design with more than seven design points so that the random errors can be precisely estimated. In this situation, one must also consider obtaining replicated observations from each design point.

Figure 8.7-2b is the diagram of simplex-lattice with four ingredients; the design coordinates for this would be a simple extension of the three ingredients simplex-lattice. The diagrams in Fig. 8.7-2c and 8.7-2d are still other configurations of the design that should provide more observations to estimate higher order coefficients if needed. Table 8.7-2 shows the simplex design coordinates for Fig. 8.7-2c. It is advantageous to include additional points not specified by the simplex to provide better prediction accuracy. The location of these points is in the interior of the simplex as specified in Table 8.7-3. Example 8.7-1 illustrates the design and analysis of a simplex-lattice experiment.

Table 8.7-2
Design coordinates for a 10-point simplex-lattice design.

Design Point	X_1	X_2	X_3
1	1	0	0
2	0	1	0
3	0	0	1
4	2/3	1/3	0
5	1/3	2/3	0
6	0	2/3	1/3
7	0	1/3	2/3
8	1/3	0	2/3
9	2/3	0	1/3
10	1/3	1/3	1/3

Table 8.7-3
Location of the interior points for the simplex-lattice design.

X_1	X_2	. . .	X_q
(q+1)/2q	1/2q	. . .	1/2q
1/2q	(q+1)/2q	. . .	1/2q
. . .	. . .	. . .	. . .
. . .	. . .	. . .	. . .
1/2q	1/2q	. . .	(q+1)/2q

Example 8.7-1

Consider an experiment on skin moisturizing lotion where palmetic acid (X_1), petrolatum (X_2), and olive oil (X_3) were studied for their effects on the perceived dryness of the skin. A 8-inch unstructured rating scale was used with anchors of no dryness (0) on one end and high dryness (8) on the other end. Table 8.7-4 displays the average scores from 24 panelists obtained in two sessions. The six formulations were evaluated following a hand dryness evaluation procedure. The fitted equation was found to be

$$Y = 5.60X_1 + 6.30X_2 + 6.40X_3 - 2.60X_1X_2 - 2.80X_1X_3 + 0.20X_2X_3$$

The SAS program used to obtain the resultant equation is as follows:

```
DATA;
INPUT X1 X2 X3 Y;
CARDS;
  1.0   0.0   0.0   5.6
  0.0   1.0   0.0   6.3
  0.0   0.0   1.0   6.4
  0.5   0.5   0.0   5.3
  0.0   0.5   0.5   6.4
  0.5   0.0   0.5   5.3
PROC GLM;
MODEL Y = X1 X2 X3 X1*X2 X1*X3 X2*X3 / NOINT;
OUTPUT OUT = POUT PREDICTED = PY1 R=RESIDUAL;
TITLE 'SKIN LOTION STUDY';
PROC PRINT;
```

When there is more than one dependent variable, the model statement becomes MODEL Y1 Y2 Y3 . . = X1 X2 X3 etc. The equation for Y can be mapped using SYSTAT/SYGRAPH or a complete statistical analysis and contour mapping using DESIGN-EXPERT.

Table 8.7-4
Design and data for the skin moisturizing lotion experiment.

Design Point	X_1	X_2	X_3	Scores	Average
1	1	0	0	5.1, 6.0	5.6
2	0	1	0	6.5, 6.0	6.3
3	0	0	1	6.7, 6.1	6.4
4	0.5	0.5	0	5.0, 5.5	5.3
5	0	0.5	0.5	6.0, 6.8	6.4
6	0.5	0	0.5	4.9, 5.6	5.3

Scale: 0 = no dryness, 8 = high dryness

Let us focus our attention to the parameter estimates of the equation. Note that the estimates b_1, b_2, and b_3 are, respectively, equal to the average scores of the pure components X_1, X_2, and X_3. Since b_3 is the largest value among the pure components, this indicates that olive oil provides the largest drying effects on the skin, followed by petrolatum with $b_2 = 6.30$, and palmetic acid with $b_1 = 5.60$. The binary blending effects b_{12} and b_{13} have negative synergism (antagonistic effect) indicating that when in combination they are perceived to result in lower scores, i.e., low dryness. Note that these results only illustrate statistical analysis and should not be experimentally interpreted.

What is the best level combination, if any, among the three moisturizing agents? A simple procedure is found by examination of the contour diagram in Fig. 8.7-3 generated by DESIGN-EXPERT. In this figure, we find that one potential combination indicated by a filled circle consists of $X_1 = 0.55$, $X_2 = 0.25$, and $X_3 = 0.20$. The predicted value of Y for this combination is obtained by substituting the ingredient levels into the fitted equation as follows:

$$\begin{aligned} Y &= 5.60(0.55) + 6.30(0.25) + 6.40(0.20) - 2.60(0.55)(0.25) \\ &\quad - 2.80(0.55)(0.20) + 0.20(0.25)(0.20) \\ &= 5.28 \end{aligned}$$

As expected, this is close to the contour line for $Y = 5.3$ on the map. The same result can be obtained for other combinations on this line.

One of the several excellent features of DESIGN-EXPERT is shown in Fig. 8.7-4. It provides the optimum operating window (low dryness score) as well as the undesirable window (high dryness score) where sensory analysts can obtain optimal formulas for use in confirmation runs. To facilitate reading this map, a triangular grid was drawn into it. Notice that the optimal area is shaded and labeled R1__LO.

Referring back to the example, the three ingredients constitute 25% of the total blend. Hence an appropriate adjustment of the ingredient level should be made for the final blend. For example, for combination with the predicted response of 5.28, the final blend is:

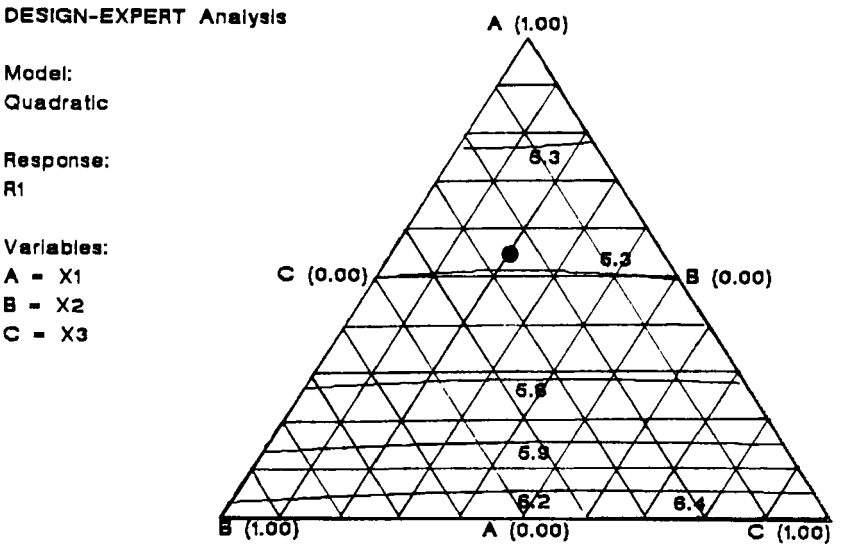


Fig. 8.7-3
Contour lines for the lotion experiment generated by DESIGN-EXPERT. Note the decreasing value of the response with increasing level of ingredient X1(A). A triangular grid is drawn on the map to facilitate reading the coordinates of a point.

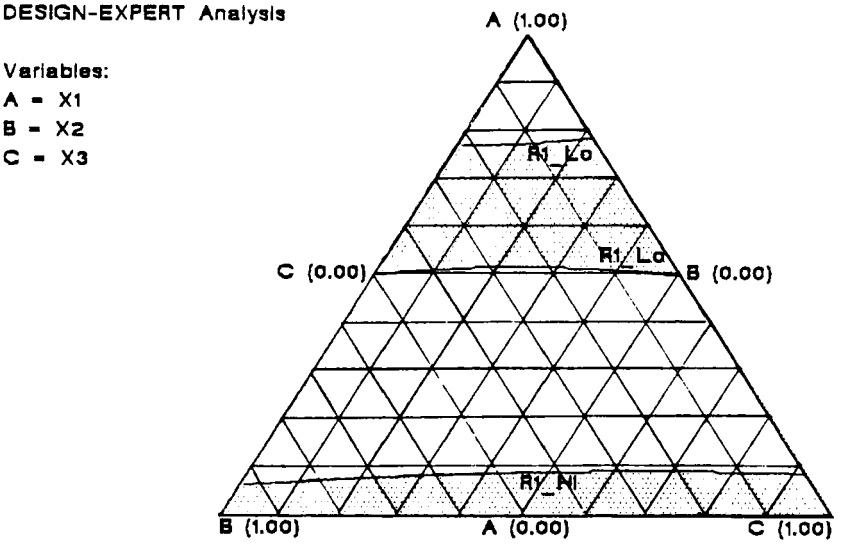


Fig. 8.7-4
Optimization map for the lotion experiment showing the operating windows (shaded) for the low (R1__LO) and high (R1__HI) values of the response.

$X_1 = 0.55(0.25)$	$= 0.1375$
$X_2 = 0.25(0.25)$	$= 0.0625$
$X_3 = 0.20(0.25)$	$= 0.0500$
Others	$= 0.7500$
Total	1.0000

Note that the adjustment did not change the proportions between the ingredients, i.e., $X_2/X_1 = 0.25/0.55 = 0.0625/0.1375 = 1/2.2$. The variance of the predicted response Y can be obtained as illustrated by Cornell (1981) in his book and in other published papers (Scheffé 1958, 1963; Gorman and Hinman 1962).

Scheffé Simplex-Centroid Design

The simplex-centroid design developed by Scheffé (1963) consists of $2^q - 1$ points in the simplex in which the ingredients are present in equal proportion. The design coordinates consist of 1, 0, .., 0 for the q pure components, 1/2, 1/2, 0, .., 0 for the binary mixtures, 1/3, 1/3, 0, .., 0 for the ternary mixtures, and 1/q, 1/q, 0, .., 0 for the q-nary mixtures (centroid).

An example of a configuration for simplex-centroid involving three and four ingredients is shown in Fig. 8.7-5 and their corresponding design coordinates given in Tables 8.7-5 and 8.7-6. As with the previous mixture designs, one may replicate each design point in the simplex to obtain measures of lack of fit and pure error. One can also augment the design in Table 8.7-5 by adding three more design points as follows (Kurotori 1966; Snee 1971):

Point	X_1	X_2	X_3
8	2/3	1/6	1/6
9	1/6	2/3	1/6
10	1/6	1/6	2/6

The first seven design points will be used to estimate the coefficients for the special cubic model and the last three points to estimate lack of fit. The location of the design points in the mixture space is given in Fig. 8.7-6.

In sensory evaluation, the observation on each design point is usually represented by the average score of several panelists.

Example 8.7-2

Three ingredients denoted by X_1, X_2, and X_3, that differ considerably in price, are used to enhance the mildness of a liquid product used for bathing and showering. These ingredients comprised 6% of the total mixture.

Using the design in Table 8.7-5, data were collected on each design point, with four replications obtained at the centroid. The data are shown in Table 8.7-7. The

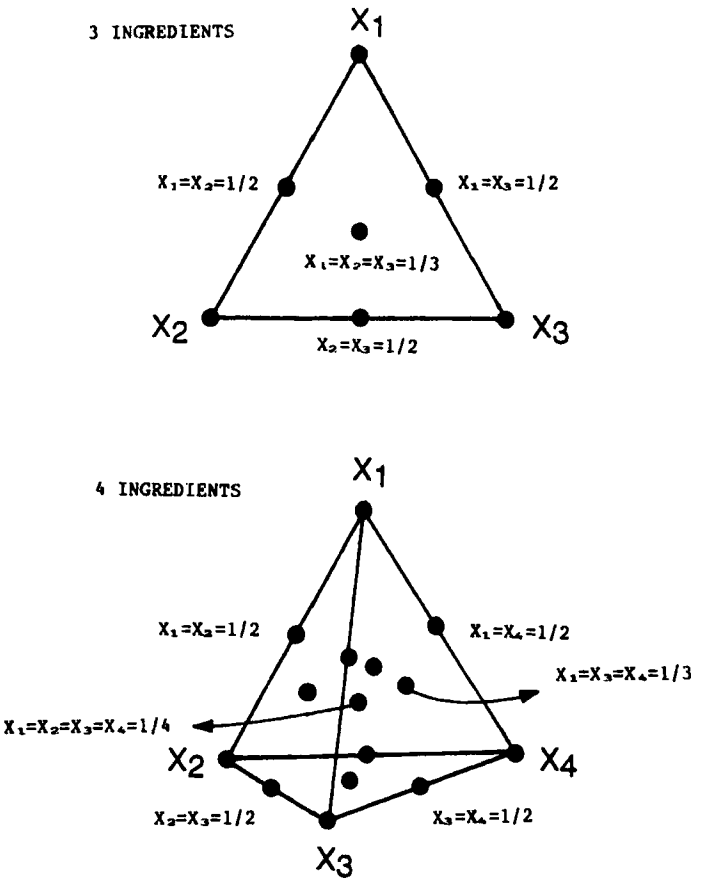


Fig. 8.7-5
Mixture space configuration of simplex-centroid design for three and four ingredients.

Table 8.7-5
Design coordinates for a 3-ingredient simplex-centroid.

Design Points	X_1	X_2	X_3
1	1	0	0
2	0	1	0
3	0	0	1
4	1/2	1/2	0
5	0	1/2	1/2
6	1/2	0	1/2
7	1/3	1/3	1/3

Table 8.7-6
Design coordinates for a 4-ingredient simplex-centroid.

Design Points	X_1	X_2	X_3	X_4
1	1	0	0	0
2	0	1	0	0
3	0	0	1	0
4	0	0	0	1
5	1/2	1/2	0	0
6	0	1/2	1/2	0
7	0	0	1/2	1/2
8	1/2	0	0	1/2
9	0	1/2	0	1/2
10	1/2	0	1/2	0
11	1/3	1/3	1/3	0
12	0	1/3	1/3	1/3
13	1/3	1/3	0	1/3
14	1/3	0	1/3	1/3
15	1/4	1/4	1/4	1/4

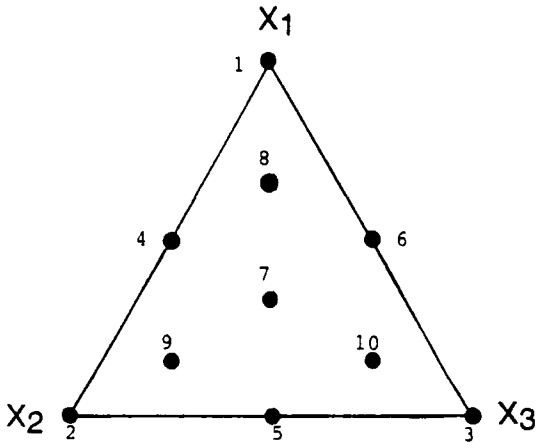


Fig. 8.7-6
Location of the design coordinates for an augmented simplex-centroid.

observations on all the design points were fitted to a quadratic model using the SAS program below:

Table 8.7-7
Experimental data for Example 8.7-2.

Design Point	X_1	X_2	X_3	Average Score
1	1	0	0	2.6
2	0	1	0	4.0
3	0	0	1	1.9
4	1/2	1/2	0	4.5
5	0	1/2	1/2	3.1
6	1/2	0	1/2	4.2
7	1/3	1/3	1/3	2.5
	1/3	1/3	1/3	3.2
	1/3	1/3	1/3	3.0
	1/3	1/3	1/3	2.8

Note: Rating scale based on a 7-point intensity scale where 1 = low mildness and 7 = high mildness.

```
DATA;
INPUT X1 X2 X3 Y;
CARDS;
1.0     0.0     0.0     2.6
0.0     1.0     0.0     4.0
0.0     0.0     1.0     1.9
0.5     0.5     0.0     4.5
0.0     0.5     0.5     3.1
0.5     0.0     0.5     4.2
0.33    0.33    0.33    2.5
0.33    0.33    0.33    3.2
0.33    0.33    0.33    3.0
0.33    0.33    0.33    2.8
PROC GLM;
MODEL Y = X1 X2 X3 X1*X2 X1*X3 X2*X3 / NOINT;
OUTPUT OUT = POUT PREDICTED = PY1 R=RESIDUAL;
TITLE 'EXAMPLE 8.7- 2';
PROC PRINT;
```

The result is shown in Table 8.7-8. The estimates of parameters of the regression equation show that only the linear effects of the ingredients were found significant. Ingredient X_2 was found to provide the largest positive effect on mildness, followed by X_1 and X_3. Although not significant, notice the antagonistic effect of X_2 and X_3. The amount of this effect is $(.5)(.5)(-2.87) = -0.72$ unit on the rating scale. In this calculation, the value 0.5 is the design coordinate for X_2 and X_3.

Table 8.7-8
Analysis of variance for Example 8.7-2 using SAS.

Source of variance		DF	SS	Parameter estimates
Linear:	b_1	1	59.65	2.77±.78*
	b_2	1	35.84	4.17±.78*
	b_3	1	7.78	2.07±.78*
Quadratic:	b_{12}	1	0.12	1.333±.37
	b_{13}	1	0.91	4.323±.37
	b_{23}	1	0.45	−2.873±.37
Residual		4	2.46	
Total		10	107.21	

* $p < 0.05$
Note: Parameter estimates ± standard error. The mean square error (MSE) is 2.46/4 = 0.62.

For the purpose of illustration, we investigate whether the model is a good fit by breaking the SSE into two parts, one due to lack of fit (SSFIT) and the other due to pure error (SSPURE) as discussed in Section 8.1. The pure error can be estimated from the observations at the centroid. This is found to be

$$\begin{aligned} \text{SSPURE} &= (2.5^2 + 3.2^2 + 3.0^2 + 2.8^2) - [(11.5^2)/4] \\ &= 33.33 - 33.06 = 0.27 \end{aligned}$$

with 4 − 1 = 3 degrees of freedom. The lack of fit SS is therefore

$$\begin{aligned} \text{SSFIT} &= \text{SSE} - \text{SSPURE} \\ &= 2.46 - 0.27 = 2.19 \end{aligned}$$

with 1 DF. Note that SSE is the residual SS in Table 8.7-8. The F-ratio, with 3 numerator DF and 1 denominator DF, is

$$\text{F-ratio} = (2.19/1)/(0.27/3) = 24.33$$

which is not significant indicating that the quadratic model provides a good fit of the mixture response surface. In fact, the adjusted R-square is $1 - \{[9(0.62)]/107.21\} = 0.95$. Since the model is a good fit, we can use the fitted equation for prediction purposes such as contour mapping, the result of which is shown in Fig. 8.7-7 and the optimization analysis shown in Fig. 8.7-8. The optimum area is located on the lower left corner of the map indicated by the shaded area. In this example, X_2 is the most expensive ingredient, and one must examine the map concentrating on the

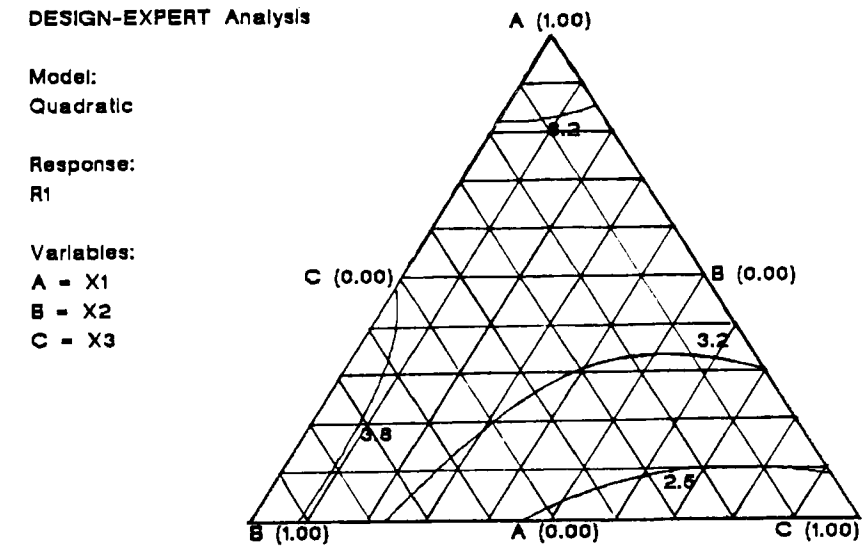


Fig. 8.7-7
Contour lines for Example 8.7-2.

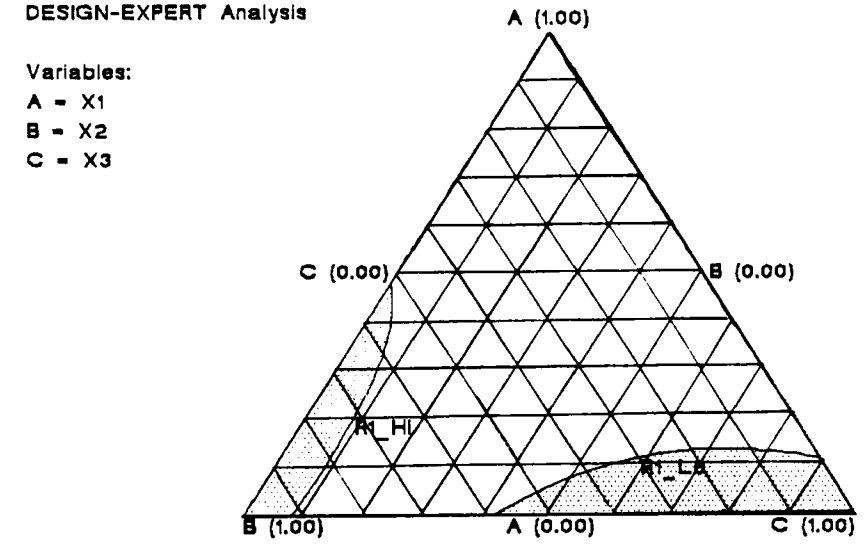


Fig. 8.7-8
Optimization map for Example 8.7-8.

combination involving X_1 and X_3. The synergistic effect of these ingredients on mildness is (0.5)(0.5)(4.33) = 1.08. The strategy is thus to find a combination with high amounts of X_1 and X_3 as a partial substitute for X2. This can be found in Fig. 8.7-8.

DESIGNS WITH CONSTRAINTS ON PROPORTION

In some situations, it is necessary to limit the levels of ingredients in the blend. Some may have only a lower limit or an upper limit, and others require both the lower and upper limits. In these situations, we are working with a subregion of the simplex. In the technical literature, these situations are known as constrained mixture experimentation. Note that in the previous sections, we have discussed mixture designs where the proportions of each ingredient are not bounded hence the entire mixture space is explored. In contrast to the constraint given by Eq. (8.7-1), the constraint now is

$$0 \leqslant a_i \leqslant X_i \leqslant b_i \leqslant 1, \qquad i = 1, 2, .., q \tag{8.7-2}$$

where a_i is the lower limit of ingredient i and b_i is its upper limit.

Lower Constraint on Proportions

We begin with a case where the ingredients have lower limits given by

$$a_i \leqslant X_i \leqslant 1$$

i.e., the proportion X_i can range from a_i to 1. Scheffé (1958) introduced the idea of pseudocomponents to facilitate formulation of mixture experiments with lower limits. Pseudocomponent is not a pure component, but is a mixture of several components, although they do not appear to be a mixture as seen by their design coordinates. The relationship between the pure components X_i, i = 1, 2, .., q, and the pseudocomponents X_i^*, is given by a simple linear transformation

$$X_i = a_i + (1 - L)X_i^* \tag{8.7-3}$$

where $L = \Sigma a_i$, the sum of lower limits for all ingredients in the mixture. From Eq. (8.7-3), the relationship below also holds:

$$X_i^* = (X_i - a_i)/(1 - L) \tag{8.7-4}$$

Example 8.7-3

An example that has been used several times in the past (Snee 1971, Cornell 1982) to illustrate a mixture experiment with lower limit on proportions is that reported by Kurotori (1966). In this experiment pertaining to rocket propellant, three components — binder, oxidizer, and fuel — are blended to find a propellant with a modulus of elasticity of 3000. The lower bounds specified in the experiment are as follows:

$$X_1 \text{ (binder)} \geqslant 0.20$$
$$X_2 \text{ (oxidizer)} \geqslant 0.40$$
$$X_3 \text{ (fuel)} \geqslant 0.20$$

Table 8.7-9 shows the design coordinates in terms of pseudocomponents and pure components along with elasticity responses for each mixture. Note that the first seven points constitute a simplex-centroid and are augmented by the addition of points 11, 12, and 13. Kurotori (1966) used the added design points to estimate lack of fit of the model when the pure error is known. For the purpose of illustration, we have modified the problem so that we can estimate the pure error. This was accomplished by replicating the centroid four times (r = 4) given by points 7, 8, 9, and 10. To reduce the size of elasticity values, the responses were divided by 100, i.e., 2350/100 = 23.50. This should not affect the results of the statistical analysis.

The pure components in Table 8.7-9 were obtained using formula (8.7-3). For example, for design point 1

$$X_1 = 0.20 + (1 - 0.80)1 = 0.40$$
$$X_2 = 0.40 + (1 - 0.80)0 = 0.40$$
$$X_3 = 0.20 + (1 - 0.80)0 = 0.20$$

and so on. Note that it is the pure component coordinates that are used to make the blend. Using the SAS program given in the previous example, let us fit the data to a special cubic model using all the 13 observations. The result is

$$Y = 23.51X_1 + 24.46X_2 + 26.53X_3 - 0.04X1X_2 + 10.10X_1X_3 + 15.99X_2X_3 + 59.47X_1X_2X_3$$

The standard error of estimates for parameters of the model are found to be

$$b_1 = b_2 = b_3 = 0.14$$
$$b_{12} = b_{13} = b_{23} = 0.69$$
$$b_{123} = 3.60$$

Table 8.7-9
Design coordinates and elasticity responses for Example 8.7-3 on rocket propellant (Kurotori 1966, with modification).

Design point	Pseudocomponents X_1	X_2	X_3	Pure components X_1	X_2	X_3	Elasticity Y
1	1	0	0	0.400	0.400	0.200	23.50
2	0	1	0	0.200	0.600	0.200	24.50
3	0	0	1	0.200	0.400	0.400	26.50
4	1/2	1/2	0	0.300	0.500	0.200	24.00
5	1/2	0	1/2	0.300	0.400	0.300	27.50
6	0	1/2	1/2	0.200	0.500	0.300	29.50
7	1/3	1/3	1/3	0.266	0.466	0.266	30.00
8	1/3	1/3	1/3	0.266	0.466	0.266	29.70
9	1/3	1/3	1/3	0.266	0.466	0.266	29.90
10	1/3	1/3	1/3	0.266	0.466	0.266	30.10
11	2/3	1/6	1/6	0.333	0.433	0.233	26.90
12	1/6	2/3	1/6	0.233	0.533	0.233	27.70
13	1/6	1/6	2/3	0.233	0.433	0.333	29.80

Note: Large values of elasticity are desirable. Values coded by Y/100.

These estimates are used to test for the statistical significance of the regression coefficients, i.e., $t = 23.51/0.14 = 167.93$, which is significant at the 0.0001 level. The SAS output shows that, except for b_{12}, which is not statistically significant, all the regression coefficients were significant ($p = .0001$).

The residual sum of squares is found to be SSE = 0.1194 with 6 degrees of freedom (Table 8.7-10). As done in the previous example, the SSE can be divided into two components, one due to lack of fit, and the other due to pure error, each with 3 degrees of freedom. Using the observations at the centroid,

$$\text{SSPURE} = (30.00^2 + 29.70^2 + 29.90^2 + 30.10^2) - [(119.70)^2/4] = 0.0875$$

and

$$\text{SSFIT} = \text{SSE} - \text{SSPURE} = 0.1194 - 0.0875 = 0.0319.$$

Then the F ratio statistic is

$$F = \text{MSFIT/MSPURE} = (0.0319/3)/(0.0875/3) = 0.3630$$

which is clearly not statistically significant (Table 8.7-10). Therefore, the special

Table 8.7-10
Analysis of variance of the rocket propellant data.

Source	DF	SS	MS	F-ratio
Pure effect				
b_1	1	5568.03		
b_2	1	2632.44		
b_3	1	1764.03		
Binary effect				
b_{12}	1	4.74		
b_{13}	1	17.35		
b_{23}	1	25.90		
Cubic effect				
b_{123}	1	5.69		
Residual	6	0.1194		
Lack of fit	3	0.0319	0.0106	0.3630
Pure error	3	0.0875	0.0292	
Total	13	1018.30		

Note: 0.1194 = 0.0319 + 0.0875

The total DF is equal to the total number of observations because the intercept b_0 was not estimated as discussed in the previous section (no intercept model).

cubic model is adequate to describe the data, and one may proceed to find the optimum combinations of the independent variables that will yield maximum elasticity.

The contour map of elasticity is shown in Fig. 8.7-9. The map portrays a mountain with maximum elasticity predicted at 0.20 binder, 0.30 oxidizer, and 0.50 fuel. In terms of pure components, a transformation is necessary using Eq. (8.7-3):

$$\text{Binder} = 0.20 + (1 - 0.80)0.20 = 0.24$$
$$\text{Oxidizer} = 0.40 + (1 - 0.80)0.30 = 0.46$$
$$\text{Fuel} = 0.20 + (1 - 0.80)0.50 = 0.30$$

Thus, the optimum combination of these components that should give a maximum elasticity is 24% binder, 46% oxidizer, and 30% fuel. As indicated in Section 8.1, this combination should be confirmed by an independent experiment. In many instances, refinements of the optimum formula are needed to meet manufacturing requirements.

Lower and Upper Constraints on Proportions

In practice we often encounter a situation where the ingredients must have lower and upper bounds to enable the product to be produced. This constraint on bounds

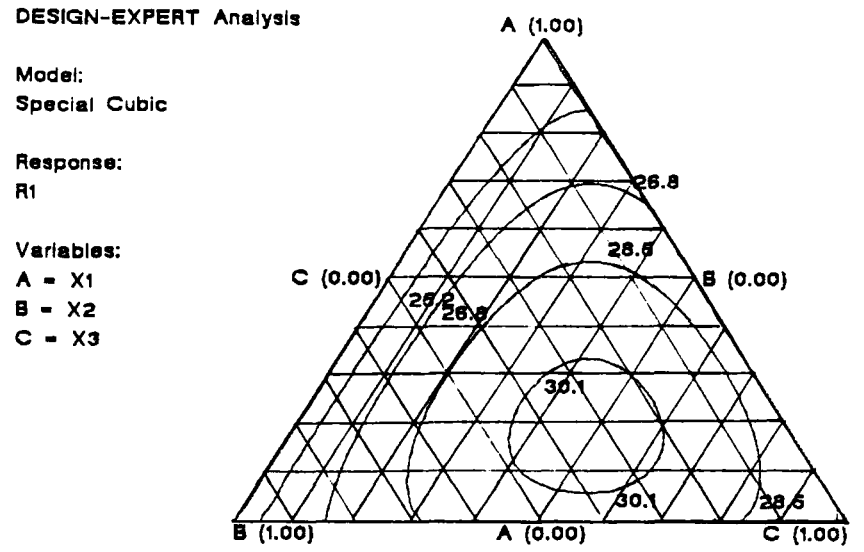


Fig. 8.7-9
Contour map of elasticity responses for the rocket propellant data (Example 8.7-3).

reduces the size of the mixture space to a subregion of the simplex creating a design problem. Solutions to this problem have been reported by McLean and Anderson (1966), Snee and Marquardt (1974), and Saxena and Nigam (1977). Let us consider the solution by McLean and Anderson, known in the literature as the extreme vertices design.

The extreme vertices design describes an irregular hyper-polyhedron with unique sets of vertices and centroids to estimate the parameters of the response surface. A minimum of $q(q + 1)/2$ points are needed to estimate the parameters of the quadratic model. Additional points may be necessary if an estimate of error and/or lack of fit is to be tested. The additional points may come from midpoints of the edges of the hyper-polyhedron or replications of existing design points. There are two steps for obtaining extreme vertices design.

Step 1. Write down all possible two level treatment combinations using the a_i and b_i levels for all but one factor, which is left blank, i.e., (a_1, b_2, a_3, ____, a_5, b_6) for a six factor (ingredient) study. This procedure generates $q(2^{q-1})$ possible treatment combinations (ingredient or component combinations) with one factor's level blank in each.

Step 2. Go through all $q(2^{q-1})$ possible treatment combinations and fill in those blanks that are admissible, i.e., that level (necessarily falling within the constraints of the missing factor), which will make the sum of the levels for that treatment com-

bination equal to one. Each of the admissible treatment combinations is a vertex; however, some vertices may appear more than once.

These steps are illustrated by the example given below.

Example 8.7-4

Consider a 3-ingredient problem (q = 3) where the ingredients are denoted by A, B, and C. Then $q(2^{q-1}) = 3(22) = 12$ points. The 12 points are the vertices (which appear more than once) of the extreme vertices design. Denote by a minus sign (−) if the ingredient is at the lower bound and a plus sign (+) if the ingredient is at the upper bound. Thus for any two-ingredient combination, one point may contain all proportions at the lower bound (−, −), another point where all proportions are at the upper bound (+, +), and two more points in the combination (+, −) and (−, +).

Suppose that the following constraints are placed on A, B, and C.

	Lower	Upper
A	0.20	0.60
B	0.12	0.50
C	0.10	0.30

Table 8.7-11 shows the method of constructing the extreme vertices design for our example. For point 1, with (−, −), this corresponds to a proportion of 0.20 for A and 0.12 for B and 1 − 0.20 − 0.12 = 0.68 for C. Since the upper bound of C is 0.60, this point is not admissible. For point 2, with (+, −), this corresponds to 0.60 for A, 0.12 for B and 1 − 0.60 − 0.12 = 0.28. This point is admissible because all proportions are within the constraints.

Of the 12 points, four points were not admissible, three were duplicates (points 3, 7, 12), and points 2, 8, and 11 clustered to each other. The clustered points can be averaged to represent a vertex. Fig. 8.7-10 shows the space configuration of the design in which four vertices are apparent. For simplicity, one can take points 6, 10, 3, and 8 as the vertices of the design. To complete the design, one can examine the configuration and locate the midpoints of the faces and a centroid. To avoid confusion with the previous design points numbering, the midpoints are indicated by points 13, 14, 15, 16 and the centroid by 17. For this example, we have a total of nine design points to explore the subregion of the simplex. The final extreme vertices design is shown in Table 8.7-12.

8.8 SEARCH FOR OPTIMUM AREAS IN RESPONSE SURFACES

In Example 8.6-1 of this chapter, maps of response surfaces for two sensory attributes were evaluated by examination of the magnitude of response for each con-

Table 8.7-11
Construction of an extreme vertices design for three ingredients.

	Factorial Combination			Corresponding Proportion		
Point	A	B	C	A	B	C
1	−	−	−	0.20	0.12	0.68*
2	+	−	·	0.60	0.12	0.28
3	−	+	·	0.20	0.50	0.30
4	+	+	·	0.60	0.50	−0.10*
5	−	·	−	0.20	0.70*	0.10
6	+	·	−	0.60	0.30	0.10
7	−	·	+	0.20	0.50	0.30
8	+	·	+	0.60	0.10	0.30
9	·	−	−	0.78*	0.12	0.10
10	·	+	−	0.40	0.50	0.10
11	·	−	+	0.58	0.12	0.30
12	·	+	+	0.20	0.50	0.30

Note: "·" = blank; * = not admissible, outside the simplex.
Constraints: $0.20 \leqslant A \leqslant 0.60$
$0.12 \leqslant B \leqslant 0.50$
$0.10 \leqslant C \leqslant 0.30$

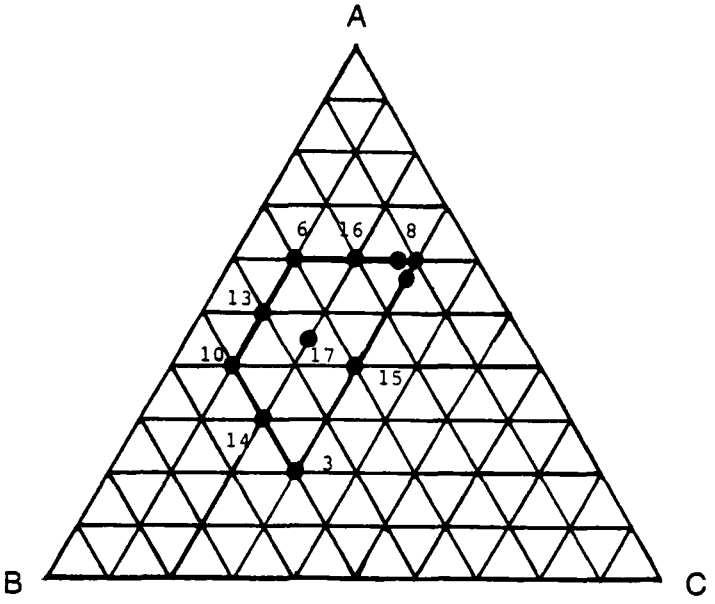


Fig. 8.7-10
Configuration of extreme vertices design in the subregion of a simplex (Example 8.7-4).

Table 8.7-12
Extreme vertices design for three ingredients with lower and upper bounds (Example 8.7-4).

Design point	A	B	C
Vertices			
6	0.60	0.30	0.10
10	0.40	0.50	0.10
3	0.20	0.50	0.30
8	0.60	0.10	0.30
Midpoints			
13	0.50	0.40	0.10
14	0.30	0.50	0.20
15	0.40	0.30	0.30
16	0.60	0.20	0.20
Centroid			
17	0.45	0.35	0.20

Note: By definition, point 17 is not a true centroid because it is not equidistant from the four vertices.

tour line on the maps. The responses for the two attributes were later linearly combined, weighted by their respective variances to simplify the search for optimum areas. When the attributes cannot be combined because of a dissimilar map of response surfaces, one can overlay the maps of the two attributes. In fact, one can overlay the maps of several attributes as illustrated in Gacula and Singh (1984) and in Box and Draper (1987). The intersection point of two or more optimal contour lines defines the coordinates of the optimum level for each ingredient.

Consider the contour maps in Fig. 8.8-1 and 8.8-2 generated by SAS using the procedure PROC RSP. The value of the response for each contour line indicated by letters A, B, C, etc., are shown at the bottom of the map. High values of Y_1 are desirable; hence, one should use contour lines E and F with predicted responses of 110 and 120, respectively (Fig. 8.8-1). Figure 8.8-2 shows another response variable Y_2. Both the dependent variables depict entirely different response surfaces; thus it is not advisable to obtain a linear combination of them. The Y_1 response depicts a "saddle" and Y_2, a "basin." However, one can overlay contour lines E and F into the surface of Fig. 8.8-2 and obtain a surface given in Fig. 8.8-3. The three filled circles are intersection of contour lines E and D, C and F, and C and E. These intersection points provide the coordinates of possible optimal levels of the independent variables. For example, point 1 is expected to give $Y_1 = 110$ and $Y_2 = 90$, point 2 to give $Y_1 = 120$ and $Y_2 = 80$, and so on. The variance of the predicted values is generally difficult to obtain and is not given here.

Another technique for searching potential areas (operating window) that is often overlooked in practice is the examination of the average responses for each design point. In some cases, large residual can be an indicator of a maximum or a minimum

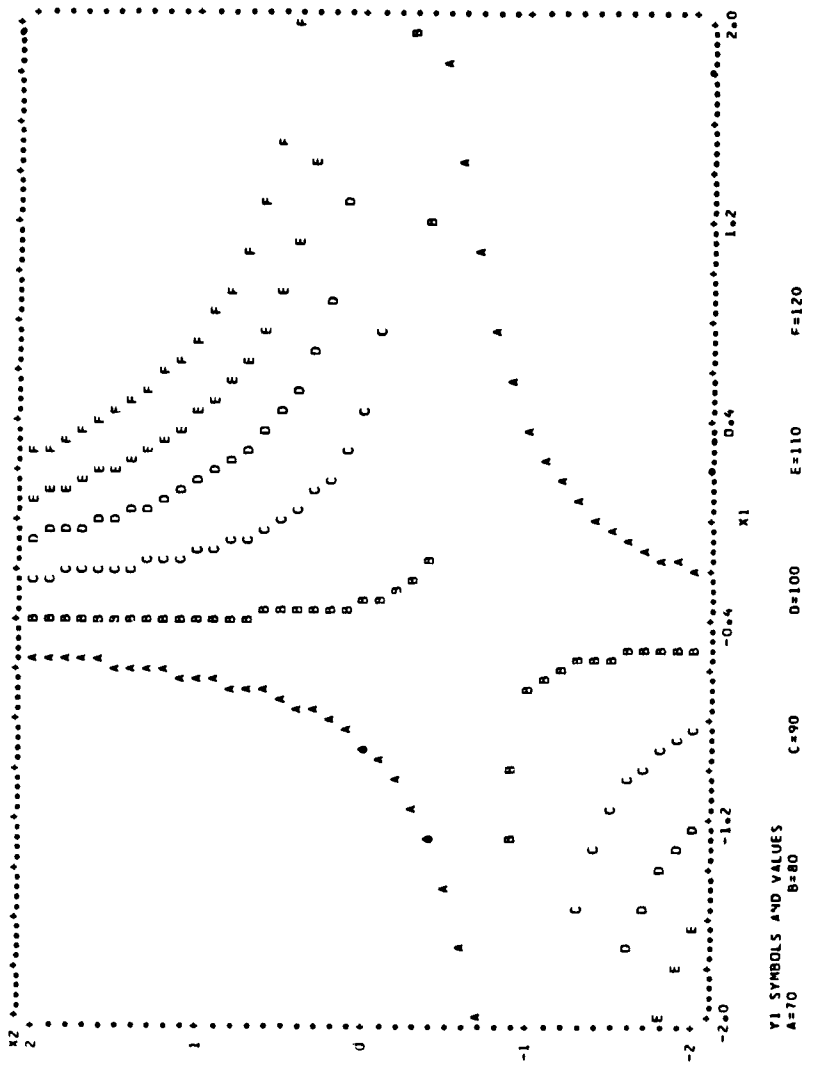


Fig. 8.8-1
Contour map for response variable Y_1.

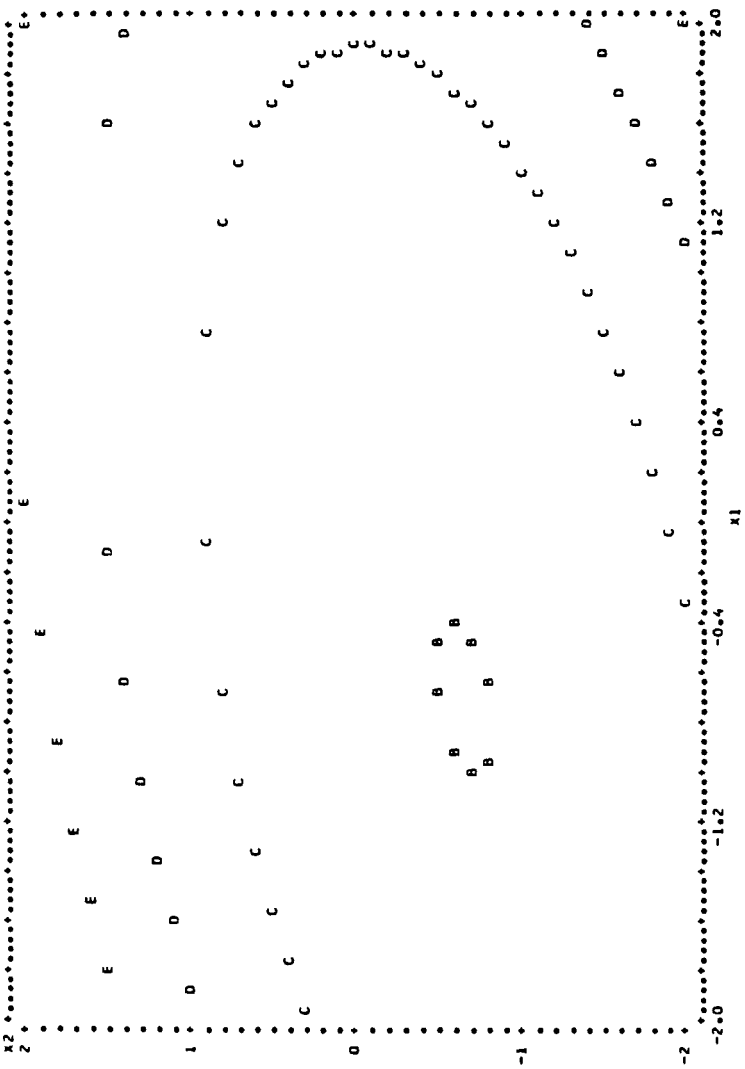


Fig. 8.8-2
Contour map for response variable Y_2.

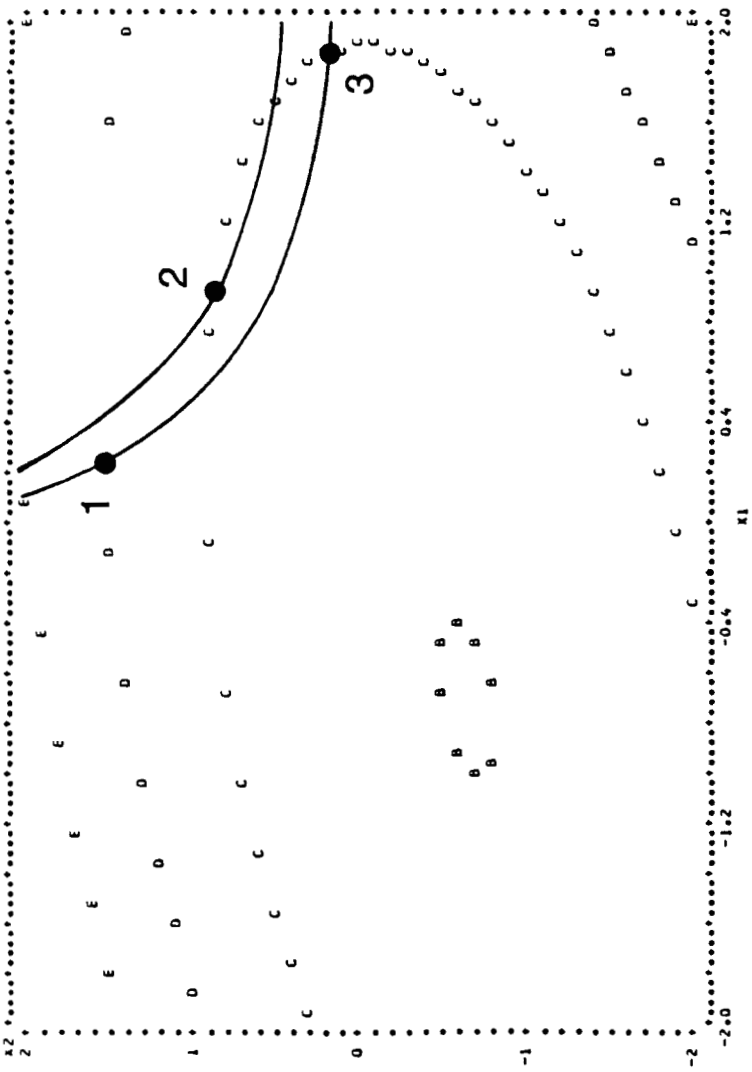


Fig. 8.8-3
Contour line overlays of response variables Y_1 and Y_2.

response; however, it is not a consistent indicator. Note that this technique should be used only for a statistically designed optimization study. In the examination of the responses, design points with desirable scores are generally good candidates for an optimum/minimum response, provided the model is a good fit as shown by a high R-square statistic. In the context of optimization, a desirable score would be high if the optimal response is the high value on the rating scale, i.e., the 9-point hedonic scale (1 = dislike extremely, 9 = like extremely). Likewise, if the optimal response is the low value on the scale, then this value is the desirable score, i.e., the 7-point off-flavor scale (0 = no off-flavor, 7 = extremely high off-flavor). In the following example, we will illustrate this technique using DESIGN-EXPERT.

Example 8.8-1

Table 8.8-1 shows the design matrix for X_1 and X_2 and dependent variables Y_1 and Y_2 from a central composite design. The estimates of the regression parameters for a quadratic response surface model are shown in Table 8.8-2, and the estimates of residuals in Table 8.8-3.

For both texture and overall liking, design point 2 with high level of X_1 and low level of X_2 has the highest average score. This signals that the operating window would be located around this area. The contour maps for Y1 (Fig. 8.8-4a) and Y2 (Fig. 8.8-4b) show that design point 2 indeed lies within the optimum operating window. If we overlay these maps, DESIGN-EXPERT will produce the optimization map shown in Fig. 8.8-5. Clearly, the shaded area on this map simultaneously optimizes texture and overall liking scores. The optimal levels for both the independent variables are found to be $X_1 = 0.20$ and $X_2 = -0.99$ and the value of Y1 at this point is a maximum at Y1 = 6.74. Result for Y2 should be similar.

Figure 8.8-6 illustrates one feature of DESIGN-EXPERT, the normal probability plot. In this figure, the residuals are plotted to determine whether they are normally

Table 8.8-1
Design matrix for a two-variable central composite design.

Design point	X_1	X_2	Y_1	Y_2
1	−1	−1	6.53	6.40
2	1	−1	6.82	7.00
3	−1	1	6.19	6.10
4	1	1	6.12	6.30
5	−2	0	6.09	6.00
6	2	0	5.85	5.50
7	0	−2	6.52	6.62
8	0	2	6.01	6.09
9	0	0	6.62	6.50

Note: Y1 = texture liking, Y2 = overall liking.

Table 8.8-2
Estimates of regression parameters for Y1 and Y2 using the DESIGN-EXPERT.

Parameter	Y1	Y2
Intercept	6.66 ± 0.13	6.65 ± 0.26
b_1	−0.02 ± 0.05	−0.02 ± 0.10
b_2	−0.17 ± 0.05	−0.17 ± 0.10
b_{11}	−0.17 ± 0.05	−0.22 ± 0.10
b_{22}	−0.10 ± 0.05	−0.07 ± 0.10
b_{12}	−0.09 ± 0.08	−0.10 ± 0.18

Note: ± standard error.

Table 8.8-3
Estimates of predicted and residual obtained from DESIGN-EXPERT. Other statistics in the output of this software are excluded for brevity.

Attribute	Design Pt.	Actual	Predicted	Residual
Y1	1	6.53	6.498	0.032
	2	6.82	6.634	0.186
	3	6.19	6.334	−0.144
	4	6.12	6.111	0.009
	5	6.09	6.024	0.066
	6	5.85	5.937	−0.087
	7	6.52	6.619	−0.099
	8	6.01	5.932	0.078
	9	6.62	6.661	−0.041
Y2	1	6.40	6.461	−0.061
	2	7.00	6.628	0.372
	3	6.10	6.318	−0.218
	4	6.30	6.084	0.216
	5	6.00	5.822	0.178
	6	5.50	5.755	−0.255
	7	6.62	6.737	−0.117
	8	6.09	6.050	0.040
	9	6.50	6.654	−0.154

Note: Y1 = Texture liking, Y2 = Overall liking. Residual = Actual − Predicted.

distributed. If the residuals fall closely along the straight line, they fit the normal distribution, and the data in this figure is a good fit for practical purposes. One can also plot the residual versus the predicted response to provide a visual analysis of the distribution of residuals (Fig.8.8-7). The residuals in this figure appeared to be distributed at random, suggesting a homogenous variance in the data.

If the desirable score in this example is a minimum, then design point 6 would be a potential optimal point. Obviously, other points on the optimization map can

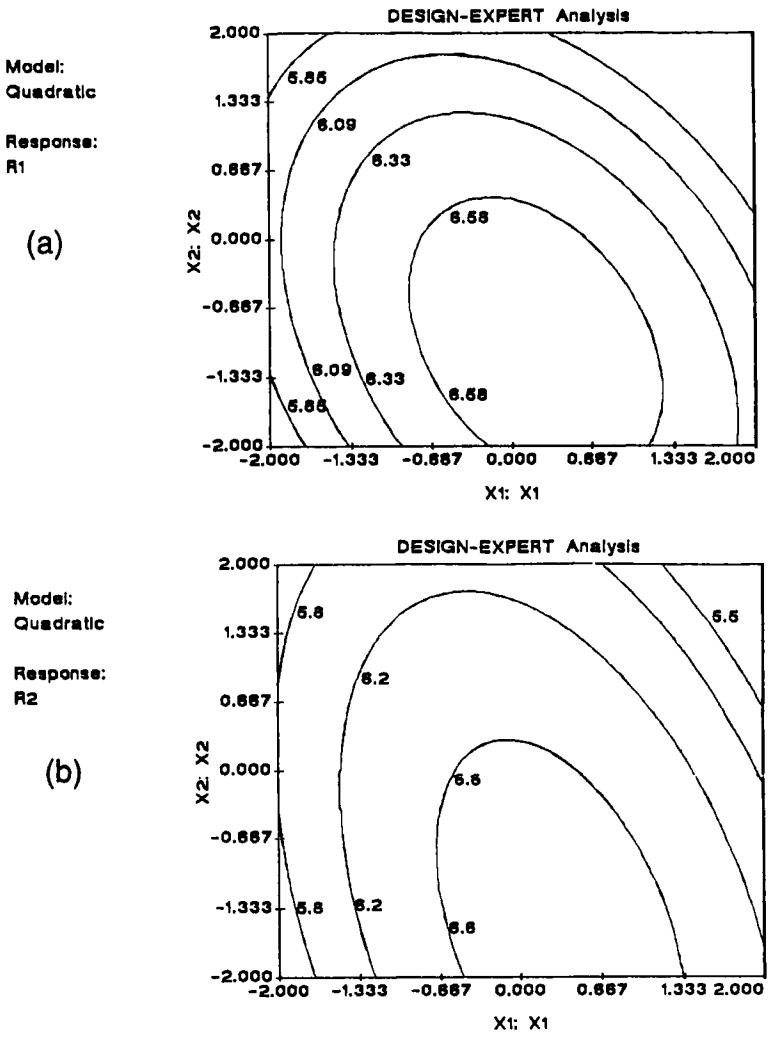


Fig. 8.8-4
Contour maps for texture (R1) and overall liking (R2) for Example 8.8-1.

be selected that meet the desirable score criterion. Again, it should be emphasized that the selected points should be validated by a confirmation run.

8.9 USE OF CONTOUR MAPS IN PRODUCT REFORMULATION

One of the many advantages of an optimization study by experimental design is that the contour maps can be used in reformulation work. In this application, the

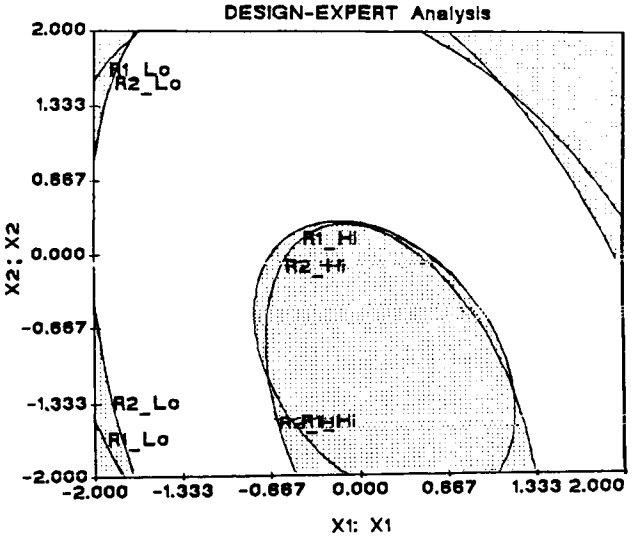


Fig. 8.8-5
Optimization map for texture (R1) and overall liking (R2) overlayed for Example 8.8-1 using the quadratic model. The low (Lo) response is 5.8 and the high (Hi), 6.60.

Fig. 8.8-6
Normal probability plot for texture liking.

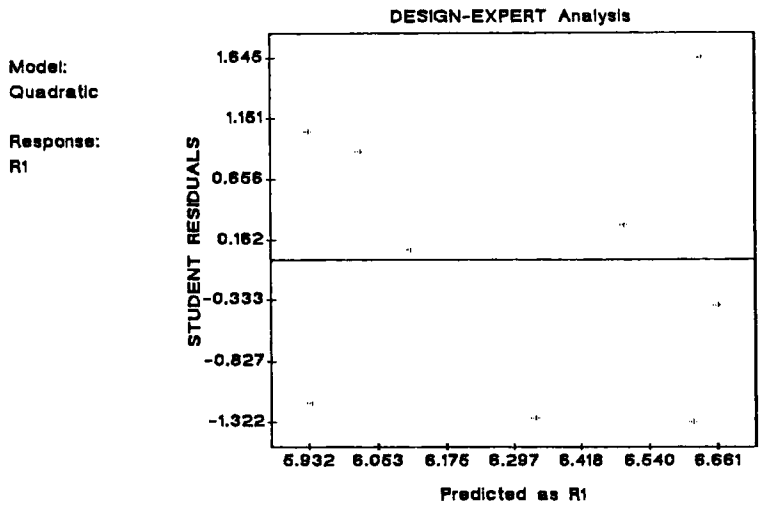


Fig. 8.8-7
Plot of residual versus predicted response for texture liking.

contour maps served as a database that contains several formulation options. Experience shows that product reformulation is a cycle that provides many challenges to individuals in research and marketing. Product reformulation occurs due to a changing market demands for certain product extensions, an increase/decrease in price of ingredients, to satisfy consumer needs, elimination of product's defects, and most importantly to be equal to or superior to competitors.

The example that follows illustrates the use of contour maps in reformulation work.

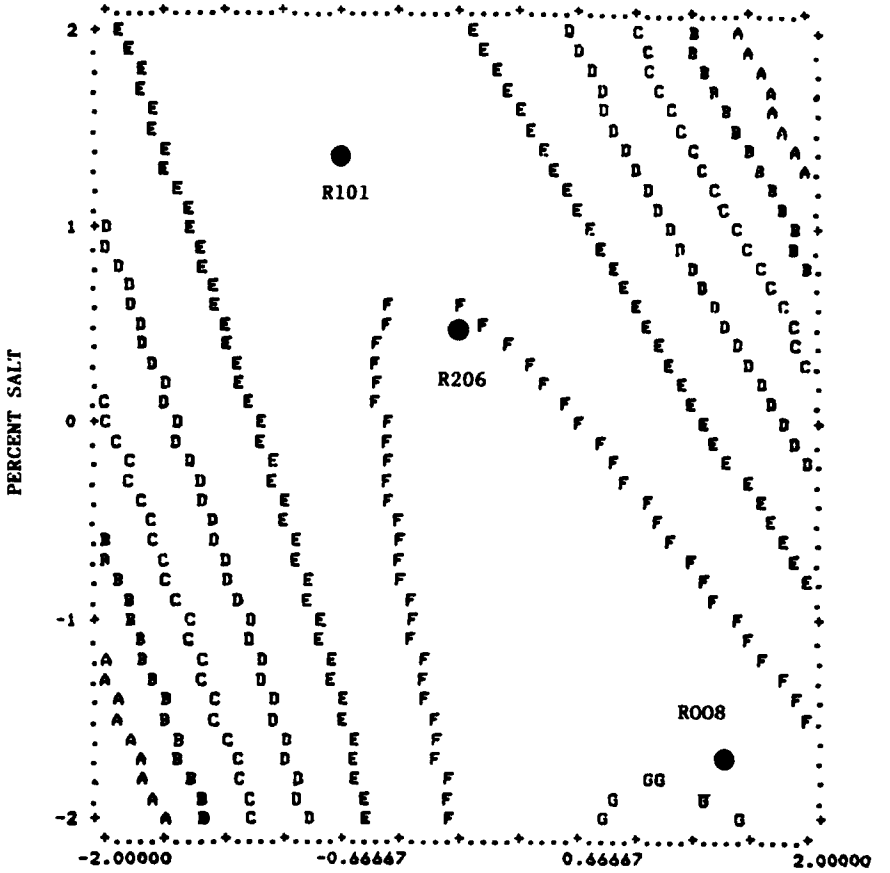
Example 8.9-1

The backgound of this study is as follows. Research and marketing personnel have focused concern on carbohydrate and salt levels of a certain canned products. Compared to their competitors, the perceived sweetness level of this product was found to be low. An optimization study was thus initiated with two variables to be varied, namely, % sweetness and % salt. The sweetness level consisted of a combination of % dextrose and % sucrose. Using a central composite design, nine formulations were made with varying levels of sweetness (0.75–2.75%) and salt (2.0–3.20%). An optimum formula was found in this study and was successfully marketed for over five years.

However, a recent audit of the product against competitors showed that the major complaint of this product is that of being bland. The initial step in approaching the problem was to review the contour maps for overall liking obtained five years ago.

This map is shown in Fig. 8.9-1. As can be seen, the map depicts a "saddle" with several potential formulas given by contour lines F and G. The marketed formula, code R008, came from the area on the lower right corner of the map. This was selected because of the consumer concern at that time of high salt level in a product.

Based on this map (Fig. 8.9-1) a reformulation was made by increasing the salt level: one formula code R206 has 2.7% salt and 1.8% sweetness and the other, code



CONTOUR LINES SYMBOLS AND VALUES

A=5.00	B=5.33	C=5.66	D=5.99
E=6.32	F=6.65	G=6.98	

Fig. 8.9-1
Contour map for overall liking. Note: R101 and R206 are reformulated formulas, R008 the current marketed formula.

R101 has 3.0% salt and 1.5% sweetness. These formulas are located in the upper left quadrant of the map. Notice that the vertical and horizontal axes are standardized, such that for % sweetness a -2 on the standardized scale corresponds to 0.75% on the actual scale and a 2 corresponds to 2.75%, and so forth; see Fig. 8.6-3.

The next step was to conduct a consumer test comparing two reformulated formulas with the current formula (R008) based on saltiness, flavor, and overall liking using the 9-point hedonic scale. The result of the consumer test shows that the formula with the 3.0% salt (code R101) was significantly more well-liked in sweetness level over the two other formulas. Based on this result, the current formula was ultimately replaced by formula code R101. In this example, it was assumed that the production processes did not change during those years.

8.10 AUGMENTATION OF FRACTIONAL FACTORIAL DESIGN

In Chapter 6, the fractional factorial design was discussed. It was shown that a considerable reduction in the number of experimental runs can be realized by fractionation. In this section, we illustrate the augmentation of fractional factorial to obtain designs for estimating quadratic response surfaces in optimization studies. Note that the Box-Wilson design presented in Section 8.6 is an augmented design accomplished by the addition of the axial and center points to the design.

Augmentation of fractional factorial is similarly obtained by the addition of the axial and center points in a manner shown below for three independent variables:

	X_1	X_2	X_3
Center point:	0	0	0
Axial points:	$-\alpha$	0	0
	α	0	0
	0	$-\alpha$	0
	0	α	0
	0	0	$-\alpha$
	0	0	α

Here, α is the length of the axial point as defined in Section 8.6 and a 0 denotes the middle level of the independent variable. The above augmentation pattern can be easily generalized to k independent variables. The Plackett-Burman and the Box-Behnken designs discussed in Sections 8.4 and 8.5, respectively, can be augmented by the above procedure. However, the statistical properties of these augmented designs are not fully understood and a theoretical work on these designs is obviously needed. In particular, because of the complex nature of the alias structure of fractional factorial, especially the Plackett-Burman design, a statistician must be consulted before using them.

Hartley (1959) reported an augmented composite design from fractional factorial to estimate quadratic response surfaces. Similarly, Westlake (1965) obtained designs based on irregular fractions of factorials and requiring even fewer experimental runs than those obtained by Hartley. In this section, we consider Hartley's design and display the design matrix to make the augmented designs useful to product formulators since the matrix indicates the levels of each ingredient that goes into the formulation of the product.

The Augmented 1/2 Fraction of 2^4

Table 8.10-1 shows the design matrix for a 1/2 fraction of 2^4 factorial design. In this design, the defining contrast I (See Section 6.2) is I = ABC which is not the usual recommendation that uses the highest order interaction ABCD. The central goal here is the estimation of response surfaces, hence we want as much as possible to estimate the two-factor interactions, in addition to the quadratic effects. This goal deviates from that discussed in Section 6.2, wherein the confounding of effects is to be minimized, i.e., minimize contamination of main effects with lower order interactions or confound higher order interactions to main effects.

Table 8.10-1
Augmented composite design for estimating quadratic response surface for a 1/2 fraction of 2^4 factorial design; I = ABC.

Design point	Factor combination	X_1	X_2	X_3	X_4
1	a	1	−1	−1	−1
2	b	−1	1	−1	−1
3	c	−1	−1	1	−1
4	abc	1	1	1	−1
5	ad	1	−1	−1	1
6	bd	−1	1	−1	1
7	cd	−1	−1	1	1
8	abcd	1	1	1	1
9	Center point	0	0	0	0
10	Axial points	$-\alpha$	0	0	0
11		α	0	0	0
12		0	$-\alpha$	0	0
13		0	α	0	0
14		0	0	$-\alpha$	0
15		0	0	α	0
16		0	0	0	$-\alpha$
17		0	0	0	α

Note: Model is $Y_{ijkm} = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + B_4X_4 + B_{14}X_1X_4 + B_{24}X_2X_4 + B_{34}X_3X_4 + B_{11}X_1^2 + B_{22}X_2^2 + B_{33}X_3^2 + B_{44}X_4^2 + E_{ijkm}$

In practice, the two-factor interaction and quadratic terms in the model are generally the most important terms that determine the curvature of the response surfaces. The design in Table 8.10-1 has the following aliases: all two-factor interactions occur in different alias-sets so that all the six interaction effects (AB, AC, AD, BC, BD, CD) can be estimated along with the four main effects (A, B, C, D) and all the quadratic effects (A^2, B^2, C^2, D^2) as shown below.

A + BC	AD + BCD
B + AC	BD + ACD
C + AB	CD + ABD
D + ABCD	

Thus, the estimate of the effects of factor A include the effects due to interaction BC, and so on. Following the above alias-sets, the statistical model of the response surface is

$$\begin{aligned} Y_{ijkm} = {} & B_0 + B_1X_1 + B_2X_2 + B_3X_3 + B_4X_4 + B_{11}X_1^2 \\ & + B_{22}X_2^2 + B_{33}X_3^2 + B_{44}X_4^2 + B_{14}X_1X_4 \\ & + B_{24}X_2X_4 + B_{34}X_3X_4 + E_{ijkm} \end{aligned}$$

Notice a change in notation to Xs, i.e., A = X_1, B = X_2; also see Section 8.1 of this chapter for the definition of the terms on the above model. In the alias-sets, three of the two-factor interactions are each confounded with main effects A, B, and C. The formulator can assign the factors accordingly, i.e., which two-factor interaction to be confounded with a particular main effect. Remember that the letters A, B, C are merely a label.

The Augmented 1/2 and 1/4 Fractions of 2^5

For the 2^5 factorial, we consider the 1/2 and the 1/4 fractions. In the 1/2 fraction, using I = ABCDE meets the appropriate alias-sets for estimating the quadratic responses and the construction of this fraction is illustrated by Example 8.10-1. For the 1/4 fraction, the recommended design matrix is given in Table 8.10-2. Here the defining contrast is I = ADE, BCE, ABCD, and all the two-factor interactions, the main and quadratic effects can be estimated. The seven alias-sets for this design are as follows:

A + DE	AB + CD
B + CE	AC + BD
C + BE	
D + AE	
E + AD + BC	

Table 8.10-2
Augmented composite design for estimating quadratic response surface for 1/4 fraction of 2^5 factorial; I = ADE, BCE, ABCD.

Design point	Factor combination	X_1	X_2	X_3	X_4	X_5
1	ab	1	1	−1	−1	−1
2	ac	1	−1	1	−1	−1
3	bd	−1	1	−1	1	−1
4	cd	−1	−1	1	1	−1
5	e	−1	−1	−1	−1	1
6	bce	−1	1	1	−1	1
7	ade	1	−1	−1	1	1
8	abcde	1	1	1	1	1
9	Center point	0	0	0	0	0
10 − 19				Axial points		

Note: Model is $Y_{ijkmn} = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + B_4X_4 + B_5X_5 + B_{11}X_1^2 + B_{22}X_2^2 + B_{33}X_3^2 + B_{44}F_4^2 + B_{55}X_5^2 + B_{12}X_1X_2 + B_{13}X_1X_3 + E_{ijkmn}$

The Augmented 1/4 Fraction of 2^6

Table 8.10-3 shows the augmented design for a 1/4 fraction of 2^6 factorial. Here the defining contrast is I = ABC, DEF, ABCDEF. As discussed in Section 6.2 (Chapter 6), the defining contrast gives the alias of an effect in the statistical model. In this table, the aliases of factor A are obtained as follows:

$$A \times ABC = BC, \quad A \times DEF = ADEF, \quad A \times ABCDEF = BCDEF$$

Therefore in this model, the estimate of effects due to factor A include effects due to interactions BC, ADEF, and BCDEF. The design in Table 8.10-3 permits the estimation of all the 15 two-factor interactions, some interactions being confounded with the six main effects, and the nine remaining aliased or confounded with three-factor or higher-order effects. For example, the two-factor alias-sets for the six main effects are:

A + BC	D + EF
B + AC	E + DF
C + AB	F + DE

The other alias-sets can be obtained as illustrated above for factor A; see also Section 6.2.

Table 8.10-3
Augmented composite design for estimating quadratic response surface for a 1/4 fraction of 2^6 factorial; I = ABC, DEF, ABCDEF.

Design point	Factor combination	X_1	X_2	X_3	X_4	X_5	X_6
1	ad	1	−1	−1	1	−1	−1
2	bd	−1	1	−1	1	−1	−1
3	cd	−1	−1	1	1	−1	−1
4	abcd	1	1	1	1	−1	−1
5	ae	1	−1	−1	−1	1	−1
6	be	−1	1	−1	−1	1	−1
7	ce	−1	−1	1	−1	1	−1
8	abce	1	1	1	−1	1	−1
9	af	1	−1	−1	−1	−1	1
10	bf	−1	1	−1	−1	−1	1
11	cf	−1	−1	1	−1	−1	1
12	abcf	1	1	1	−1	−1	1
13	adef	1	−1	−1	1	1	1
14	bdef	−1	1	−1	1	1	1
15	cdef	−1	−1	1	1	1	1
16	abcdef	1	1	1	1	1	1
17	Center point	0	0	0	0	0	0
18 – 29				Axial points			

Note: Model is $Y_{ijkmnp} = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + B_4X_4 + B_5X_5 + B_6X_6 + B_{14}X_1X_4 + B_{24}X_2X_4 + B_{34}X_3X_4 + B_{15}X_1X_5 + B_{25}X_2X_5 + B_{35}X_3X_5 + B_{16}X_1X_6 + B_{26}X_2X_6 + B_{36}X_3X_6 + B_{11}X_1^2 + B_{22}X_2^2 + B_{33}X_3^2 + B_{44}X_4^2 + B_{55}X_5^2 + B_{66}X_6^2 + E_{ijkmnp}$

Example 8.10-1

In this example, the five independent variables are denoted by X_1, X_2, X_3, X_4, and X_5, and the dependent variable by Y_1. Let us follow the fractionation procedure discussed in Chapter 6. For a complete factorial with k = 5 variables there are 32 factorial combinations (Table 8.10-4). This number of combination is obviously prohibitive in many research situations, thus, we may fractionate these combinations to a half to result in two sets, each set with 16 combinations. Confounding the highest order interaction results in the defining equation

$$L = X_a + X_b + X_c + X_d + X_e$$

Applying this equation to the 32 treatment combinations divides them into 2 sets, each combination with either L mod 2 = 1 or 0. If we take L mod 2 = 0, the fractional factorial combinations are shown in Table 8.10-5. Estimate of effects for AB will include effects due to interactions AB and CDE; effect AC will include effects due to interactions AC and BDE, and so on.

Table 8.10-4
Factorial combinations for k = 5 independent variables to illustrate confounding and fractionation.

Factor combination	X_a	X_b	X_c	X_d	X_e	L	L mod 2
(1)	0	0	0	0	0	0	0
a	1	0	0	0	0	1	1
b	0	1	0	0	0	1	1
ab	1	1	0	0	0	2	0
c	0	0	1	0	0	1	1
ac	1	0	1	0	0	2	0
bc	0	1	1	0	0	2	0
abc	1	1	1	0	0	3	1
d	0	0	0	1	0	1	1
ad	1	0	0	1	0	2	0
bd	0	1	0	1	0	2	0
abd	1	1	0	1	0	3	1
cd	0	0	1	1	0	2	0
acd	1	0	1	1	0	3	1
bcd	0	1	1	1	0	3	1
abcd	1	1	1	1	0	4	0
e	0	0	0	0	1	1	1
ae	1	0	0	0	1	2	0
be	0	1	0	0	1	2	0
abe	1	1	0	0	1	3	1
ce	0	0	1	0	1	2	0
ace	1	0	1	0	1	3	1
bce	0	1	1	0	1	3	1
abce	1	1	1	0	1	4	0
de	0	0	0	1	1	2	0
ade	1	0	0	1	1	3	1
bde	0	1	0	1	1	3	1
abde	1	1	0	1	1	4	0
cde	0	0	1	1	1	3	1
acde	1	0	1	1	1	4	0
bcde	0	1	1	1	1	4	0
abcde	1	1	1	1	1	5	1

Note: Defining equation $L = X_a + X_b + X_c + X_d + X_e$

In this example, L mod 2 = 1 is used and the design matrix for this is given in Table 8.10-6. Each factorial combination (formulation) is numbered from 1 to 16 representing each point of the design. Remember that the numbers 1, −1, etc., are coordinate of each point in the design space configuration. The design matrix which consists of these coordinate points is orthogonal to simplify obtaining the regression

Table 8.10-5
Aliases and a one-half fraction of a 2^5 factorial.

Factorial combinations with L mod 2 = 0	Pairs of aliases
(1)	μ, ABCDE
ab	AB, CDE
ac	AC, BDE
bc	BC, ADE
ad	AD, BCE
bd	BD, ACE
cd	CD, ABE
abcd	ABCD, E
ae	AE, BCD
be	BE, ACD
ce	CE, ABD
abce	ABCE, D
de	DE, ABC
abde	ABDE, C
acde	ACDE, B
bcde	BCDE, A

Table 8.10-6
Design matrix for L mod 2 = 1.

Factorial combination	Point	X_1	X_2	X_3	X_4	X_5
a	1	1	−1	−1	−1	−1
b	2	−1	1	−1	−1	−1
c	3	−1	−1	1	−1	−1
abc	4	1	1	1	−1	−1
d	5	−1	−1	−1	1	−1
abd	6	1	1	−1	1	−1
acd	7	1	−1	1	1	−1
bcd	8	−1	1	1	1	−1
e	9	−1	−1	−1	−1	1
abe	10	1	1	−1	−1	1
ace	11	1	−1	1	−1	1
bce	12	−1	1	1	−1	1
ade	13	1	−1	−1	1	1
bde	14	−1	1	−1	1	1
cde	15	−1	−1	1	1	1
abcde	16	1	1	1	1	1
Center point	17	0	0	0	0	0
Axial points	18–27					

equation. In Table 8.10-6 formulation 1 (Point 1) consists of variable X_1 at the high level denoted by 1 and all the remaining variables at the low level denoted by -1, and so on.

The design in Table 8.10-6 is augmented by the addition of a center point and five axial points corresponding to each independent variable. That is, a total of 2k + 1 points is added or 11 points in our example. Thus the total number of formula-

Table 8.10-7
Design matrix, responses, and SAS program to analyze the data for Example 8.10-1.

```
DATA;
TITLE 'OPTIMIZATION STUDY OF A HOUSEHOLD PRODUCT';
INPUT X1-X5 Y1;
CARDS;
 1   -1   -1   -1   -1    8.56
-1    1   -1   -1   -1    6.72
-1   -1    1   -1   -1    7.30
 1    1    1   -1   -1   10.84
-1   -1   -1    1   -1    8.87
 1    1   -1    1   -1   12.46
 1   -1    1    1   -1   12.13
-1    1    1    1   -1    9.55
-1   -1   -1   -1    1    6.67
 1    1   -1   -1    1    9.84
 1   -1    1   -1    1   11.05
-1    1    1   -1    1    8.74
 1   -1   -1    1    1   11.33
-1    1   -1    1    1    9.01
-1   -1    1    1    1    6.91
 1    1    1    1    1   10.74
 0    0    0    0    0   10.28
-2    0    0    0    0    5.65
 2    0    0    0    0   11.43
 0   -2    0    0    0    9.33
 0    2    0    0    0   10.40
 0    0   -2    0    0    9.31
 0    0    2    0    0   10.33
 0    0    0   -2    0    6.47
 0    0    0    2    0   10.97
 0    0    0    0   -2   10.00
 0    0    0    0    2   10.21
PROC GLM;
MODEL Y1 = X1 X2 X3 X4 X5 X1*X1 X2*X2 X3*X3 X4*X4 X5*X5
           X1*X2 X1*X3 X2*X3 X1*X4 X2*X4 X3*X4 X1*X5 X2*X5
           X3*X5 X4*X5;
OUTPUT OUT = POUT PREDICTED = PY1 RESIDUAL = RES;
PROC PRINT;
```

tions for a 5-variable study is 16 + 11 = 27. The resulting design matrix along with the value of each response for each design point is shown with the SAS program in Table 8.10-7. Note that point 17 represents the center point of the design where all the five variables are set at the middle level. Also, the length of the axial point is $\alpha = \pm 2$. The estimates of parameters of the model as will as the predicted responses are given in Table 8.10-8. Notice that in the parameter estimates, the main effects are confounded with the 4-factor interactions and the 2-factor interactions are confounded with the 3-factor interactions. These confounding patterns meet our goal to estimate the curvature of the response surface. In this example, a high response value is desirable and an inspection of the predicted values (PY1) reveals some interesting formulations. For example, the following formulations (Obs) are potentially promising:

Formula	Predicted Response
6	12.46
7	12.18
11	10.94
13	11.49

Contour maps around these formulations should be explored.

In a well-designed experiment, one usually finds one or more design points closely meeting the criteria of response and the selected point can be used as a preliminary formula. In this case, a considerable amount of time and money is saved in the optimization process.

8.11 PRECAUTION ON FRACTIONAL FACTORIAL DESIGNS

We have discussed in Chapter 6 and in the previous section how to reduce the number of design points or treatment combinations to minimize cost of experimentation. It was indicated that this reduction results in loss of information through confounding of unimportant effects. There is no way, other than conducting an experiment, to verify that the assumed unimportant effects are in fact truly unimportant, since they are confounded to other effects in the model. This assumption is the major precaution in the use of fractional factorial designs, particularly in screening studies, i.e., estimation of main (factor) effects. To minimize the consequence of this assumption, one must use a fractional design where higher order interaction effects, i.e., 3-factor interaction, are confounded to the main effects. The software DESIGN-EASE provides a more complete tabulation of fractional factorial as well as the full factorial designs. In this section, the above assumption will be examined using sensory data from an optimization study.

Table 8.10-8
A partial SAS output for Example 8.10-1.

OPTIMIZATION OF A HOUSEHOLD PRODUCT

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y1

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	10.07777778	24.32	0.0001	0.41441850
X1	1.44750000	15.09	0.0001	0.09591924
X2	0.30083333	3.14	0.0202	0.09591924
X3	0.24333333	2.54	0.0443	0.09591924
X4	0.84500000	8.81	0.0001	0.09591924
X5	-0.07166667	-0.75	0.4832	0.09591924
X1*X1	-0.35916667	-3.06	0.0223	0.11747660
X2*X2	-0.02791667	-0.24	0.8201	0.11747660
X3*X3	-0.03916667	-0.33	0.7502	0.11747660
X4*X4	-0.31416667	-2.67	0.0368	0.11747660
X5*X5	0.03208333	0.27	0.7939	0.11747660
X1*X2	-0.21625000	-1.84	0.1153	0.11747660
X1*X3	0.08375000	0.71	0.5027	0.11747660
X1*X4	0.09125000	0.78	0.4668	0.11747660
X1*X5	0.00500000	0.04	0.9674	0.11747660
X2*X3	-0.00750000	-0.06	0.9512	0.11747660
X2*X4	-0.00250000	-0.02	0.9837	0.11747660
X2*X5	-0.02125000	-0.18	0.8624	0.11747660
X3*X4	-0.53000000	-4.51	0.0041	0.11747660
X3*X5	-0.16375000	-1.39	0.2128	0.11747660
X4*X5	-0.49375000	-4.20	0.0057	0.11747660

OPTIMIZATION OF A HOUSEHOLD PRODUCT

OBS	X1	X2	X3	X4	X5	Y1	PY1	RES
1	1	-1	-1	-1	-1	8.56	8.3169	0.24306
2	-1	1	-1	-1	-1	6.72	6.4461	0.27389
3	-1	-1	1	-1	-1	7.30	7.0711	0.22889
4	1	1	1	-1	-1	10.84	10.5753	0.26472
5	-1	-1	-1	1	-1	8.87	8.9094	-0.03944
6	1	1	-1	1	-1	12.46	12.4636	-0.00361
7	1	-1	1	1	-1	12.13	12.1786	-0.04861
8	-1	1	1	1	-1	9.55	9.5678	-0.01778
9	-1	-1	-1	-1	1	6.67	6.5536	0.11639
10	1	1	-1	-1	1	9.84	9.6878	0.15222
11	1	-1	1	-1	1	11.05	10.9428	0.10722
12	-1	1	1	-1	1	8.74	8.6019	0.13806
13	1	-1	-1	1	1	11.33	11.4911	-0.16111
14	-1	1	-1	1	1	9.01	9.1403	-0.13028
15	-1	-1	1	1	1	6.91	7.0853	-0.17528
16	1	1	1	1	1	10.74	10.8794	-0.13944
17	0	0	0	0	0	10.28	10.0778	0.20222
18	-2	0	0	0	0	5.65	5.7461	-0.09611
19	2	0	0	0	0	11.43	11.5361	-0.10611
20	0	-2	0	0	0	9.33	9.3644	-0.03444
21	0	2	0	0	0	10.40	10.5678	-0.16778
22	0	0	-2	0	0	9.31	9.4344	-0.12444
23	0	0	2	0	0	10.33	10.4078	-0.07778
24	0	0	0	-2	0	6.47	7.1311	-0.66111
25	0	0	0	2	0	10.97	10.5111	0.45889
26	0	0	0	0	-2	10.00	10.3494	-0.34944
27	0	0	0	0	2	10.21	10.0628	0.14722

Using the confounding and fractional techniques given in Chapter 6, it can be shown that the following two sets of treatment combinations result from a one-half fraction of 2^3 using the ABC interaction as the defining contrast:

Set 1	Set 2
a	(1)
b	ab
c	ac
abc	bc

In Set 1, factor A is confounded with the BC interaction, B with the AC interaction, and C with the AB interaction. Note that the ABC interaction is confounded with the grand mean μ. The confounding of effects is easily seen by considering the design matrix below:

	μ	A	B	AB	C	AC	BC	ABC
a	+	+	−	−	−	−	+	+
b	+	−	+	−	−	+	−	+
c	+	−	−	+	+	−	−	+
abc	+	+	+	+	+	+	+	+

For simplicity, the coefficient of 1 is omitted. The heading of this matrix (first row) are effects to be estimated if a full factorial (nonfractional) is used. Notice that the coefficients for μ and ABC are identical since they are confounded; likewise, for A and BC, and so on. If one uses Set 2, the result will be the same except for signs. Thus, $A = -BC$, $B = -AC$, and $C = -AB$ and the sums of squares of these effects will also be the same.

Let us change notation by using X_1 for A, X_2 for B and X_3 for factor C. If Set 1 is used in the optimization experiment, the model will be

$$Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + \text{error}$$

However, because of confounding the expected values (E) of the regression coefficients are

$$E(B_1) = B_1 + B_{23}$$
$$E(B_2) = B_2 + B_{13}$$
$$E(B_3) = B_3 + B_{12}$$

That is, the regression coefficients for the main effects are confounded with the interaction regression coefficients. Unless theoretical and/or historical data are available, one must conduct Set 2 experiment to verify the magnitude of the interaction regression coefficients.

Example 8.11-1

We can evaluate fractional factorial designs from data obtained from a full factorial experiment since each treatment combination is independent of the others. Let us use the data from a smoked ham optimization with three independent variables X_1, X_2, and X_3. In this example, the dependent variable Y is overall liking measured on a 150-mm unstructured line scale with larger values denoting an increasing liking for the product.

A central composite design with 15 design points was used in this study. Table 8.11-1 displays the design matrix and the average sensory score Y computed from a balanced incomplete block design. Notice that design points 1–8 represent the 2^3 factorial design, point 9 the center point, and points 10–15 the axial points. In the statistical analysis, the first 8 points was split into two representing Sets 1 and 2 as discussed at the beginning of this section. Three models were fit to the data, one for the central composite design (full model), and the remaining fractional models for Set 1 and Set 2.

The equation to be solved for the unknown regression parameters for Set 1 data is:

$$\begin{matrix} B_0 & X_1 & X_2 & X_3 \end{matrix} \quad \begin{matrix} Y \end{matrix}$$

$$\begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & -2 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 1 & 0 & -2 & 0 \\ 1 & 0 & 2 & 0 \\ 1 & 0 & 0 & -2 \\ 1 & 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 90 \\ 75 \\ 72 \\ 100 \\ 92 \\ 75 \\ 99 \\ 80 \\ 90 \\ 79 \\ 80 \end{bmatrix}$$

Table 8.11-1
Design matrix and overall liking data from a central composite design optimization study.

Design point	X_1	X_2	X_3	Y
1	−1	−1	−1	70
2	1	−1	−1	90
3	−1	1	−1	75
4	1	1	−1	98
5	−1	−1	1	72
6	1	−1	1	90
7	−1	1	1	101
8	1	1	1	100
9	0	0	0	92
10	−2	0	0	75
11	2	0	0	101
12	0	−2	0	80
13	0	2	0	90
14	0	0	−2	79
15	0	0	2	80

Note: Column of 1s for B_0 is omitted.

Notice that we used design points 2, 3, 5, 8 for the factorial portion of the design matrix given in Table 8.11-1. For Set 2, design points 1, 4, 6, and 7 will be used. All the three models will use the center and the axial design points. The result for Set 1 is

$$b = (X'X)^{-1}X'Y = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_{11} \\ b_{22} \\ b_{33} \end{bmatrix} = \begin{bmatrix} 91.00 \\ 7.58 \\ 2.75 \\ 0.75 \\ -1.13 \\ -1.63 \\ -3.00 \end{bmatrix}$$

Table 8.11-2 summarizes the results for the three models. The general observation of the estimates is that they appear to be similar, all models showing that X_1 and X_2 were the major factors for overall liking of the product. Since the interaction effects were not statistically significant, the estimates of main effects are likely to be unbiased.

Table 8.11-3 shows the predicted values obtained from the three models. These predicted values were obtained by substituting the observed values into the fitted equation. The residual estimates provide information on how good is the fractional factorial in predicting the results from the full model when certain interaction ef-

Table 8.11-2
Estimates of regression parameters for the full and fractional factorial models (Set 1 and Set 2).

Parameter	Full model	Set 1	Set 2
b_0	93.00	91.00	94.38
b_1	6.75**	7.58**	5.42*
b_2	4.50**	2.75*	4.92*
b_3	2.00	0.75	2.08
b_{11}	−1.38	−1.13	−1.55
b_{22}	−1.88	−1.63	−2.05
b_{33}	−3.25	−3.00*	−3.42
b_{12}	−2.00		
b_{13}	−3.25		
b_{23}	3.25		
R-square	0.89	0.92	0.76

Note: Estimates of b_{12}, b_{13}, and b_{23} for Sets 1 and 2 are confounded with b_1, b_2, and b_3.
* $p \leq .10$
** $p < .01$

Table 8.11-3
Observed, predicted, and residual values for the three models.

Design point	Observed	Predicted			Residual		
		Full	Set 1	Set 2	Full	Set 1	Set 2
1	70	71		75	−1		−5
2	90	95	89		−5	1	
3	75	78	80		−3	−5	
4	98	94		96	4		2
5	72	75	76		−3	−4	
6	90	86		90	4		0
7	101	95		90	6		11
8	100	98	96		2	4	
9	92	93	91	94	−1	1	−2
10	75	74	71	77	1	−4	−2
11	99	101	102	99	−2	−3	0
12	80	77	79	76	3	1	4
13	90	95	90	96	−5	0	−6
14	79	76	78	77	3	1	2
15	80	84	81	85	−4	−1	−5

Note: Residual = observed − predicted.
Design point: 4 = high X_1 and X_2, low X_3
7 = high X_2 and X_3, low X_1
8 = all variables at high level
11 = extreme X_1, middle level for X_2 and X_3

fects are negligible. At this stage of the analysis, we may find potential formulations based on high mean scores such as design points 4, 7, 8, and 11. It is likely that optimal formulas may be found in these neighboring areas. These formulas can be searched in the contour maps shown in Fig. 8.11-1 for the three models. Since the effect of X_3 was not significant, it is appropriate to plot X_1 and X_2 and hold X_3 at the middle level, $X_3 = 0$.

The similarity among the contour maps in Fig. 8.11-1 points to the usefulness of fractional factorial in optimization studies. For example, the combination ($X_1 = 0.0$, $X_2 = -1.2$) gives an overall liking response of between 85 and 90 units on the scale for the full (Fig. 8.11-1a) and the fractional models (Fig. 8.11-1b, c).

8.12 OPTIMIZATION OF DISCRETE VARIABLES

It is common in product development that one deals with discrete variables such as, color of package, shape of container, type of fragrance, and many others. These classes of variables can be simultaneously included in an optimization of continuous

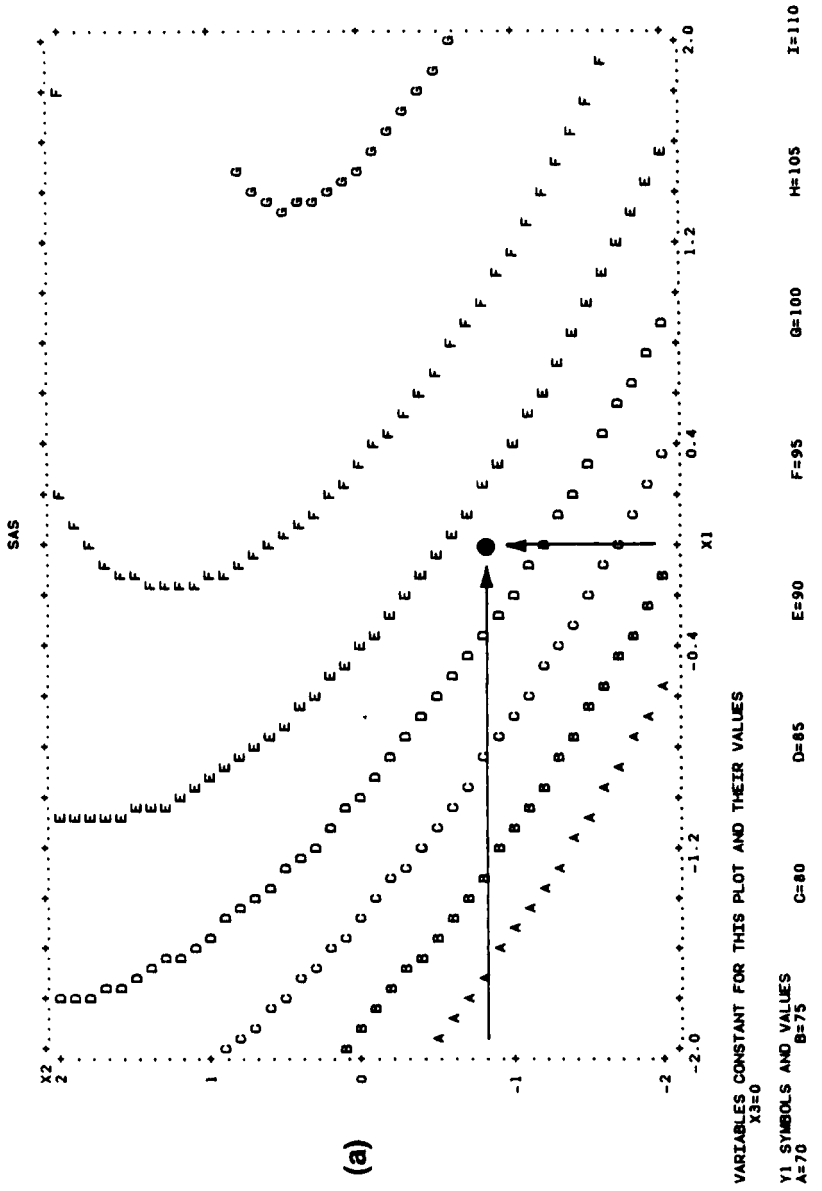


Fig. 8.11-1
Contour maps for the full and fractional factorial models. Filled circles have coordinate of (0.0, −1.2); (a) = Full model, (b) = Set 1 fractional model, (c) = Set 2 fractional model.

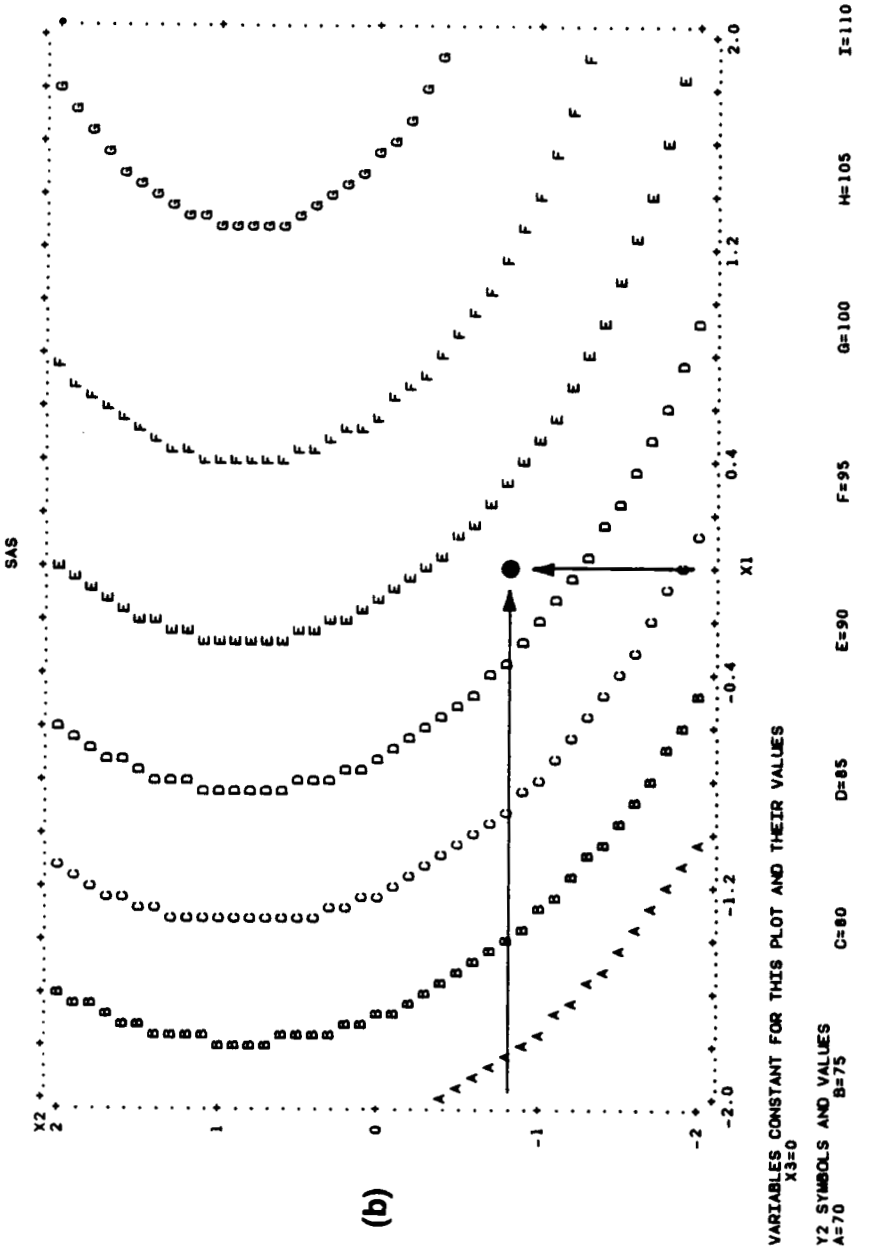


Fig. 8.11-1 (*Continued*)

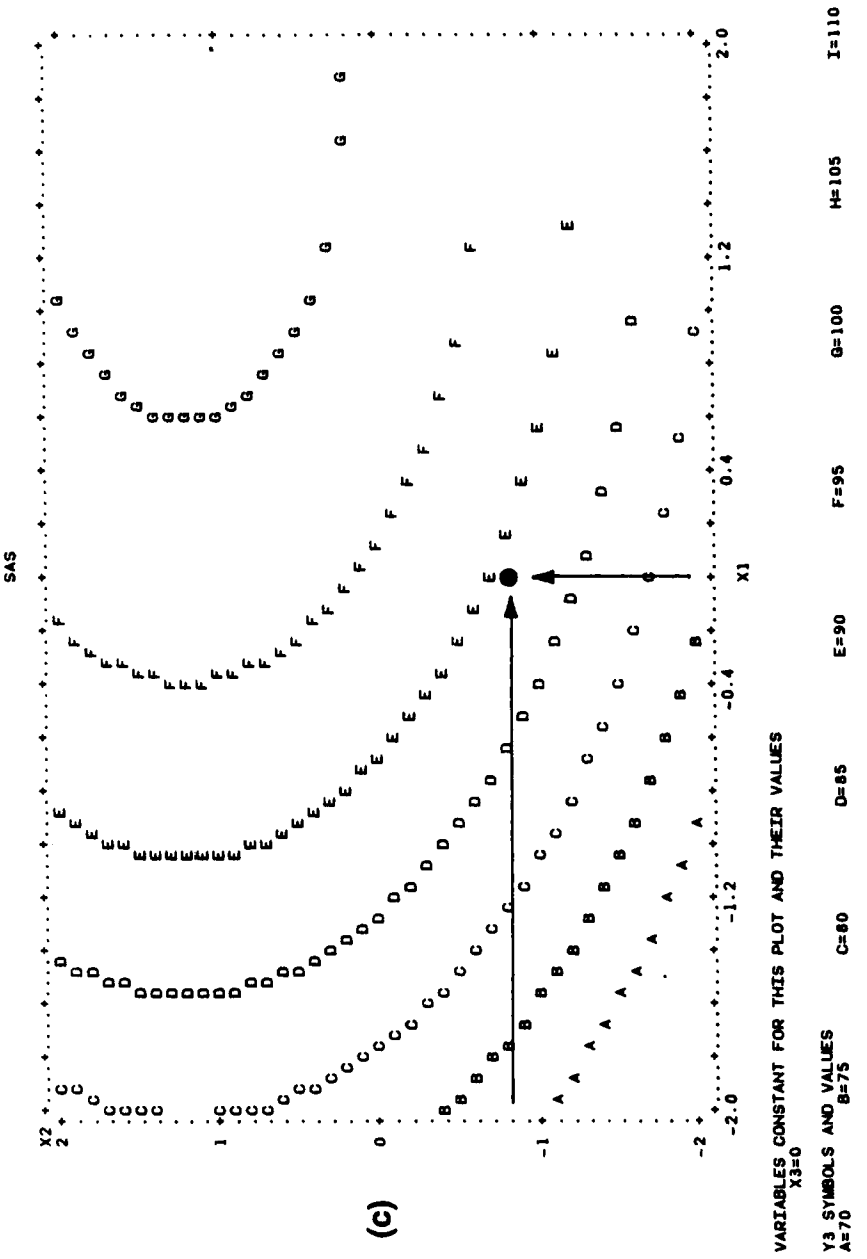


Fig. 8.11-1 (*Continued*)

variables. Note that continuous variables can assume quantitative levels. In a regression analysis terminology, discrete variables are called dummy or indicator variables.

This section illustrates the technique and data analysis of optimization study that includes both continuous and discrete variables. The method of data analysis is a direct application of multiple regression. Moskowitz (1983, 1984) presented a nonstatistical discussion of discrete variable optimization as applied to food and cosmetic products. References that discussed indicator variables include the books by Neter *et al.* (1974) and by Draper *et al.* (1981).

Discrete Variable Optimization

The optimization that we have discussed so far pertains to continuous variables, such as the levels of salt in a formulation. In other situations, one works only with discrete variables. Let us illustrate a discrete variable optimization by way of an example. This example also illustrates a technique of data analysis using mean scores from an incomplete block design to estimate interaction effects between discrete variables.

Example 8.12-1

A household product formula has been developed that meets several performance criteria. The next step in the development process is the selection of color and fragrance to be used on the finished product. It was decided to conduct a consumer test to determine the best combination of color and fragrance type as perceived by users of the product.

Two colors, blue and white, and two types of fragrances denoted by A and B are to be optimized. In an optimization work aimed at competing with existing products in the marketplace it is always recommended to include a control sample in the comparison, preferably the market leader. In this study, a 2×2 factorial design was used that gives the following color by fragrance combinations:

Formula	Color	Fragrance
1	white	A
2	white	B
3	blue	A
4	blue	B

Formula 5 was denoted the control sample. A home-use test was conducted using a balanced incomplete block design augmented with a control sample so that all the four formulas were compared in the presence of the control sample.

Table 8.12-1 shows the basic design of the home-use test. Note that with p repetitions of the basic design, the design parameters b, r_t, and λ were multiplied by p.

Table 8.12-1
Basic design for the color-fragrance study.

	Formula				Control
Panelist	1	2	3	4	5
1	x	x			x
2	x		x		x
3	x			x	x
4		x	x		x
5		x		x	x
6			x	x	x
Design parameters:	$t+1 = 5$				
	$b = 6$				
	$r_t = 3$				
	$r_c = b = 6$				
	$\lambda = 1$				

As shown in this table, each panelist evaluated two formulas plus the control during a 3-day test period, one product per day. The control sample was always used first without the knowledge of the panelists and the order of use for the remaining samples was randomized. The products were scored after each use on a 9-point hedonic scale. For the purpose of illustration, the data from three repetitions ($p = 3$) of the basic design were used. Table 8.12-2 shows the data for overall liking along with relevant calculations following the formulas given in Table 4.2-2 for the analysis of variance (Chapter 4), and likewise for the estimation of treatment effects t_i. The analysis of variance table is given in Table 8.12-3. The effects of treatments (including the control) were statistically significant at the 1% level. Multiple comparison of means is not done here because it is irrelevant in optimization. The adjusted means by the intrablock analysis for the four formulas and the control are as follows (See Chapter 4 for calculations of adjusted means):

Formulas	Adjusted Means
1 (white/A)	6.0
2 (white/B)	6.7
3 (blue/A)	5.3
4 (blue/B)	7.5
5 (control)	6.3

Based on the above means, Formula 4 is found to be the optimal combination of color and fragrance consisting of blue and the B fragrance.

Let us examine how the panelists assign relative importance of color and fragrance in their perception of these variables. This is accomplished by first analyzing the

Table 8.12-2
Overall liking data for Example 8.12-1. Design parameters: t+1 = 5, k+1 = 3, p = 3, pb = 18, pr = 9, pλ = 3

	Control	Formula					
Panelist	5	1	2	3	4	$X_{.j}$	R_m
1	6	5	6			17	
2	7	6		6		19	
3	5	6			8	19	
4	7		6	4		17	
5	6		7		8	21	
6	8			6	8	22	115
7	6	5	7			18	
8	7	6		6		19	
9	6	7			7	20	
10	7		7	6		20	
11	5		6		7	18	
12	6			5	7	18	113
13	6	6	7			19	
14	5	7		6		18	
15	6	6			7	19	
16	7		7	5		19	
17	6		6		7	19	
18	6			6	8	20	114
$X_{i.}$	112	54	59	50	67	G=342	
$B_{(i)}$	342	168	168	172	176		
$B_{(i)}/k+1$	114	56	56	57.3333	58.6667		
Q_i	−2	−2		−7.3333	8.3333	$\Sigma Q_i=0$	
t_i	−.1167	−.3571	.3571	−1.1190	1.1190		

Calculations:
CF = $(342)^2/54 = 2166.0$
SSTO = $(6^2 + 5^2 + \ldots + 6^2 + 8^2) - 2166 = 42.0$
SSR = $[(115^2 + 113^2 + 114^2)/18] - 2166 = 0.1111$
SSBL:R = $[(17^2 + 19^2 + \ldots + 19^2 + 20^2)/3] - SSR - CF = 9.2222$
SST = $(-.3571)(-2) + \ldots + (-.1167)(-2) = 19.5498$
SSE = $42.0 - 0.1111 - 9.2222 - 19.5498 - 13.1169$

data by repetition. The adjusted means obtained are used as the observations for the analysis of variance. The adjusted means are shown in Table 8.12-4. Recall that t_i are estimates of treatment (formula) effects. Using the adjusted means as the response variable Y, the analysis of variance model is

$$Y_{ijk} = \mu + C_i + F_j + (CF)_{ij} + E_{ijk}$$

Table 8.12-3
Analysis of variance for the color-fragrance study.

Source of Variance	DF	SS	MS	F-ratio
Total	53	42.0000		
Repetition	2	0.1111	0.0556	0.14
Panelist within repetition	15	9.2222	0.6148	1.50
Treatment (adjusted)	4	19.5498	4.8875	11.92*
Error	32	13.1169	0.4099	

*$p < 0.01$

Table 8.12-4
Estimates of treatment effects and adjusted means by repetition.

Repetition	Formula	Effect t_i	Adjusted Mean
I	1	−.5000	5.83
	2	.3571	6.69
	3	−1.3571	4.98
	4	1.5000	7.83
II	1	−.5000	5.83
	2	.5000	6.83
	3	−.9286	5.40
	4	.9286	7.26
III	1	−.0714	6.43
	2	.2143	6.71
	3	−1.0714	5.43
	4	.9286	7.43

where μ is the grand mean, C_i the effects of color, F_j the effects of fragrance type, $(CF)_{ij}$ the interaction effects of color and fragrance on perceived liking of the product, and E_{ijk} the random error effects. The main interest in this analysis is the estimate of interaction effects.

Table 8.12-5 shows the analysis of variance of the adjusted means in Table 8.12-4. Fragrance was found to be the driving force in overall liking ($p < 0.0001$). Color effect across fragrance types was not significant due to the significant color x fragrance interaction ($p < 0.001$), indicating that the perception of color and fragrance in a product is highly dependent of each other. Because of this interaction, the effects of color and fragrance should be evaluated simultaneously as shown in Fig. 8.12-1, i.e., with fragrance A, the white color was more well-liked than the blue color, whereas with fragrance B, the blue color was more well-liked. The adjusted mean for the control sample is 6.3, which is considerably lower than the mean of the blue/fragrance B combination of 7.5, thus this combination (Formula 4) is the choice in the optimization process.

Table 8.12-5
Analysis of variance of adjusted treatment means to estimate the interaction effects between color and fragrance.

Source of Variance	DF	SS	MS	F-ratio
Total	11	8.8166		
Fragrance	1	6.5269	6.5269	95.05*
Color	1	0.0000	0.0000	0.00
Fragrance × color	1	1.7404	1.7404	25.35*
Error	8	0.5493	0.0687	

*$p < .001$

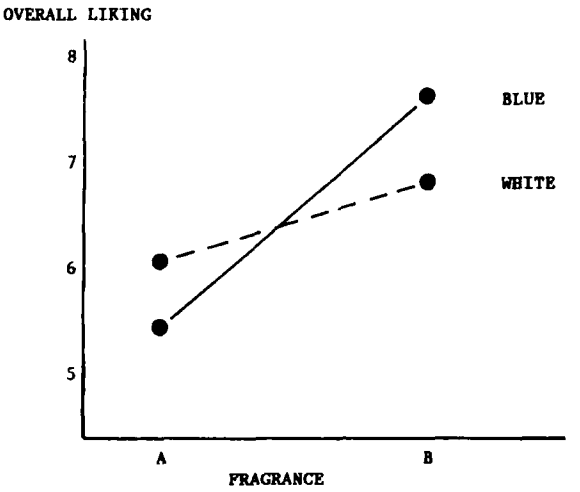


Fig. 8.12-1
A plot of the interaction between fragrance type and color of product.

Optimization of Discrete and Continuous Variables

We will now discuss a problem where the variables in the optimization include both discrete and continuous. The statistical solution is again a direct application of the multiple linear regression technique. Let us consider an example.

Example 8.12-2

A 2 × 2 factorial design with center point was conducted for a laundry aid product. The configuration of this design is shown in Fig. 8.12-2. Table 8.12-6 gives the design matrix and the response variable Y instrumentally obtained and expressed

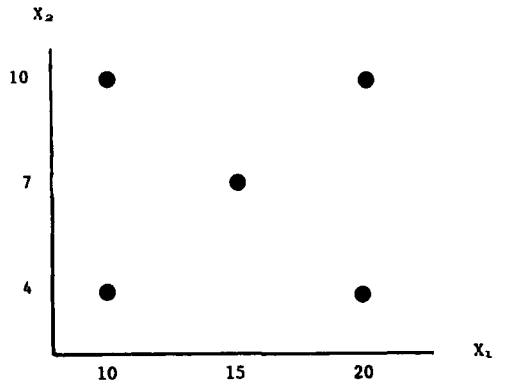


Fig. 8.12-2
A 2 × 2 factorial design with center point for Example 8.12-2.

Table 8.12-6
Design matrix for the laundry aid study to illustrate the design and analysis.

Design point	Coded Level			Actual Level			Y
	X_1	X_2	X_3	X_1	X_2	X_3	
1	−1	−1	0	10	4	0	60.0
2	1	−1	0	20	4	0	54.1
3	−1	1	0	10	10	0	71.3
4	1	1	0	20	10	0	71.7
5	0	0	0	15	7	0	64.3
1	−1	−1	1	10	4	1	61.1
2	1	−1	1	20	4	1	59.8
3	−1	1	1	10	10	1	75.7
4	1	1	1	20	10	1	77.9
5	0	0	0	15	7	1	68.6

Note: The constant column B_0 of 1's is omitted. Generally, computer package requires only the above matrix as input.

as a percentage of product performance during use. Note that the factorial design is repeated twice to accommodate the discrete variable X_3 with coefficient of 0 for container shape A and 1 for container shape B. The remaining variables, X_1 and X_2, are ingredients for enhancing the performance of the product.

The regression equation is obtained by the least squares Eq. (8.1-5), $b = (X'X)^{-1}X'Y$, where X is the design matrix in Table 8.12-6 and Y is the vector of responses. In this example, the actual level was used in the least squares analysis to avoid the problem of a singular matrix. This problem arises because the column for the discrete variable in the design matrix is not orthogonal, i.e., column does not sum to zero.

$$X = \begin{array}{c} \begin{matrix} b_0 & X_1 & X_2 & X_3 \end{matrix} \\ \begin{bmatrix} 1 & 10 & 4 & 0 \\ 1 & 20 & 4 & 0 \\ 1 & 10 & 10 & 0 \\ 1 & 20 & 10 & 0 \\ 1 & 15 & 7 & 0 \\ 1 & 10 & 4 & 1 \\ 1 & 20 & 4 & 1 \\ 1 & 10 & 10 & 1 \\ 1 & 20 & 10 & 1 \\ 1 & 15 & 7 & 1 \end{bmatrix} \end{array}$$

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 10 & 20 & 10 & 20 & 15 & 10 & 20 & 10 & 20 & 15 \\ 4 & 4 & 10 & 10 & 7 & 4 & 4 & 10 & 10 & 7 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$Y = \begin{bmatrix} 60.0 \\ 54.1 \\ 71.3 \\ 71.7 \\ 64.3 \\ 61.1 \\ 59.8 \\ 75.7 \\ 77.9 \\ 68.6 \end{bmatrix}$$

Then

$$b = (X'X)^{-1}X'Y = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_{12} \end{bmatrix} = \begin{bmatrix} 56.613 \\ -0.687 \\ 1.342 \\ 4.340 \\ 0.082 \end{bmatrix}$$

The output based on the STATISTIX package is shown in Tables 8.12-7 and 8.12-8. Note that the X_4 column is the interaction of X_1 and X_2 obtained through the TRANSFORMATION command. Table 8.12-8 contains the variance-covariance matrix obtained by the formula $\sigma^2(X'X)$. This matrix is used to compute standard errors of the regression coefficients. In the above formula, σ^2 is estimated by the error mean square in the regression analysis, which is found to be 1.581. From the variance-covariance matrix, we obtained the following standard errors:

Table 8.12-7
Least squares analysis for the data in Table 8.12-6. Note that $X_4 = X_1 \times X_2$, the interaction between X_1 and X_2.

VIEW DATA

CASE	X1	X2	X3	Y	X4
1	10.000	4.0000	0.0000	60.000	40.000
2	20.000	4.0000	0.0000	54.100	80.000
3	10.000	10.000	0.0000	71.300	100.00
4	20.000	10.000	0.0000	71.700	200.00
5	15.000	7.0000	0.0000	64.300	105.00
6	10.000	4.0000	1.0000	61.100	40.000
7	20.000	4.0000	1.0000	59.800	80.000
8	10.000	10.000	1.0000	75.700	100.00
9	20.000	10.000	1.0000	77.900	200.00
10	15.000	7.0000	1.0000	68.600	105.00

UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF Y

PREDICTOR VARIABLES	COEFFICIENT	STD ERROR	STUDENT'S T	P
CONSTANT	56.613	3.5855	15.79	0.0000
X1	-6.8667E-01	2.2572E-01	-3.04	0.0287
X2	1.3417	4.6863E-01	2.86	0.0353
X3	4.3400	7.9529E-01	5.46	0.0028
X4	8.1667E-02	2.9639E-02	2.76	0.0401

CASES INCLUDED	10	MISSING CASES	0
DEGREES OF FREEDOM	5		
OVERALL F	84.76	P VALUE	0.0001
ADJUSTED R SQUARED	0.9738		
R SQUARED	0.9855		
RESID. MEAN SQUARE	1.581		

$$SE(b_0) = \sqrt{12.86} = 3.586$$
$$SE(b_1) = \sqrt{0.0509} = 0.226$$
$$SE(b_2) = \sqrt{0.2196} = 0.469$$
$$SE(b_3) = \sqrt{0.7953} = 0.795$$
$$SE(b_{12}) = \sqrt{0.000878} = 0.030$$

See also Table 8.12-7. In Table 8.12-8, one finds the residual (RES) of each design point calculated by Y − FIT.

As shown in Table 8.12-7, the estimates of regression parameters were all significant (P column, $p < 0.05$). To evaluate the effects of container on product performance, one substitutes the value 0 or 1 into the fitted regression equation. For example, $X_3 = 0$ for container A; if we choose $X_1 = 16.0$ and $X_2 = 5.5$ as indicated in Fig. 8.12-3a, we obtain

$$Y = 56.61 - 0.687(16.0) + 1.342(5.5) + 4.340(0) + 0.082(16.0)(5.5)$$
$$= 60.2\%.$$

For container B, $X_3 = 1$, and upon substitution into the equation we obtain $Y = 64.6\%$ (Fig. 8.12-3b). Therefore, container B is to be selected because it is 4.3% more in performance than A. Since container is a discrete variable, the difference in performance between containers is its regression coefficient in the model. The contour maps in Fig. 8.12-3 also provide other choices of the leels of X_1 and X_2 that give the desired product performance. It also provides the levels that should be avoided, for they would result in poor performance. The slight curvature of the contour lines is the contribution of the interaction effects in the model.

Table 8.12-8
Variance-covariance matrix and estimates of predicted values (FIT) and residuals (RES).
Note: −7.642E-01 = −0.7642, 5.095E-02 = 0.05095, and so on.

VARIANCE - COVARIANCE MATRIX FOR COEFFICIENTS

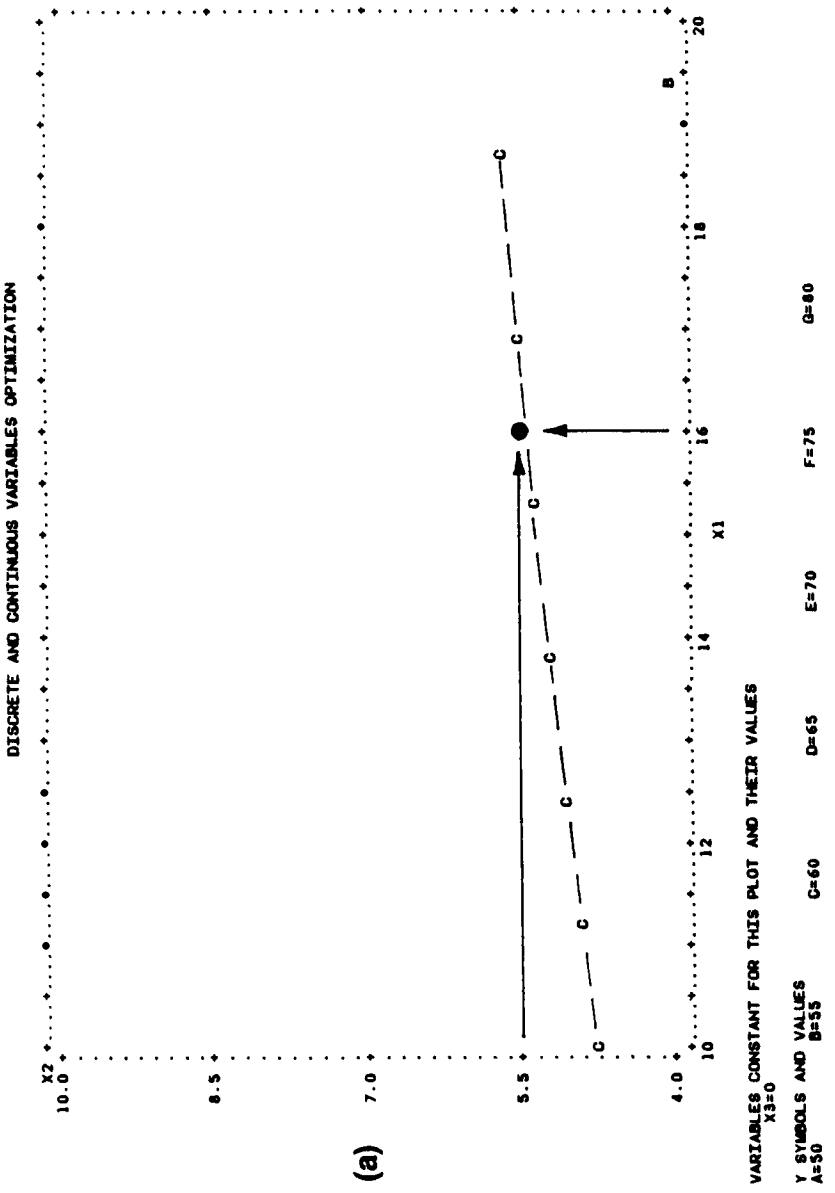
	CONSTANT	X1	X2	X3	X4
CONSTANT	12.86				
X1	-7.642E-01	5.095E-02			
X2	-1.537	9.224E-02	2.196E-01		
X3	-3.162E-01	5.560E-17	1.342E-16	6.325E-01	
X4	9.224E-02	-6.149E-03	-1.318E-02	-7.629E-18	8.784E-04

VIEW DATA

CASE	X1	X2	X3	Y	X4	RES	FIT
1	10.000	4.0000	0.0000	60.000	40.000	1.6200	58.380
2	20.000	4.0000	0.0000	54.100	80.000	-0.6800	54.780
3	10.000	10.000	0.0000	71.300	100.00	-0.0300	71.330
4	20.000	10.000	0.0000	71.700	200.00	-0.9300	72.630
5	15.000	7.0000	0.0000	64.300	105.00	0.0200	64.280
6	10.000	4.0000	1.0000	61.100	40.000	-1.6200	62.720
7	20.000	4.0000	1.0000	59.800	80.000	0.6800	59.120
8	10.000	10.000	1.0000	75.700	100.00	0.0300	75.670
9	20.000	10.000	1.0000	77.900	200.00	0.9300	76.970
10	15.000	7.0000	1.0000	68.600	105.00	-0.0200	68.620

8.13 OPTIMIZATION FOR ROBUSTNESS

We have discussed in the past sections the traditional approach to product optimization that emphasizes statistical modeling of responses and control/isolation of extraneous variations by statistical design. Because of this emphasis, we may develop products and processes that require special conditions before they can perform satisfactorily at the hands of the consumer. In other words, the products are not


DISCRETE AND CONTINUOUS VARIABLES OPTIMIZATION
X2
10.0
8.5
7.0
5.5
4.0
10
12
14
16
18
20
X1
B
C
(a)
VARIABLES CONSTANT FOR THIS PLOT AND THEIR VALUES
X3=0
Y SYMBOLS AND VALUES
A=50 B=55 C=60 D=65 E=70 F=75 G=80

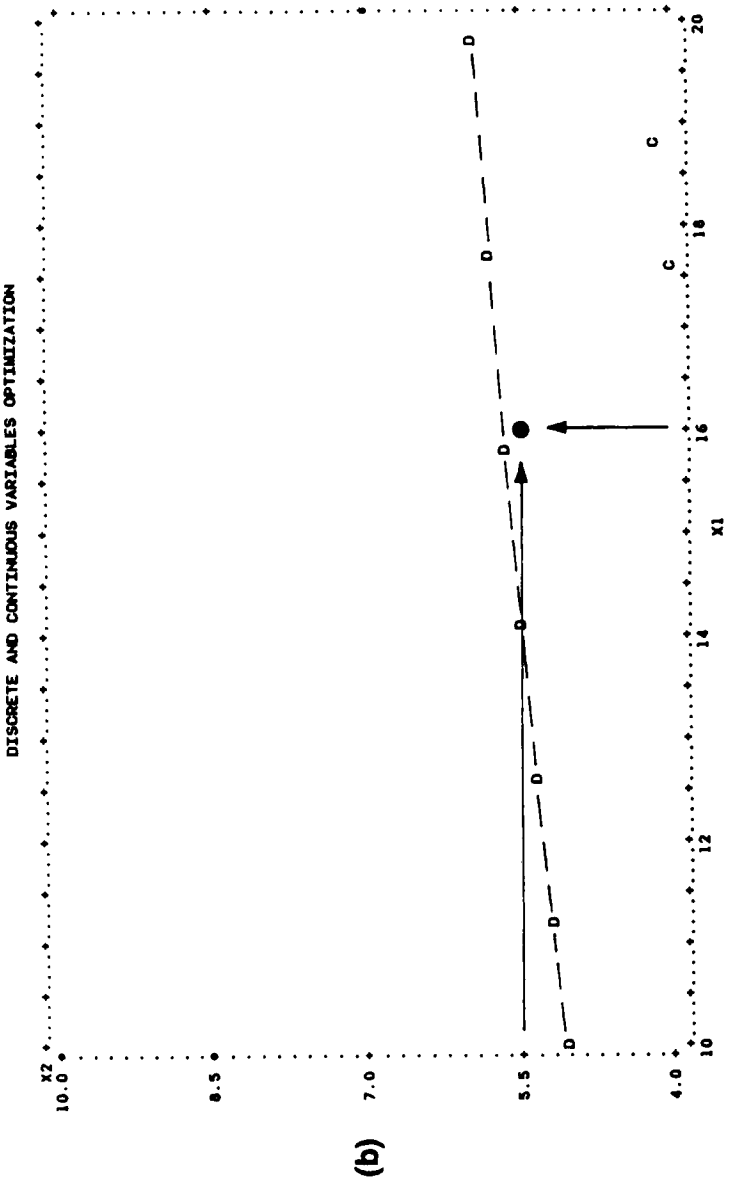


Fig. 8.12-3
Contour maps for discrete and continuous variables for Example 8.12-2.

robust and are sensitive to environmental factors during use. A robust product must be insentive to these factors, and robustness is one of the central concepts of total quality.

Dr. Genichi Taguchi pioneered a method by designing quality into products and processes, the results of which are documented in his book on quality engineering (Taguchi 1986). This method is aimed at producing robust products and manufacturing processes at low cost or without increasing cost. The success of this method in engineering and manufacturing is well-documented; however certain statistical aspects of the Taguchi methods have been criticized by statisticians (Box *et al.* 1988, Montgomery 1991a). Taguchi defined quality as the loss imparted by the product/service to the customer or society. The traditional definition includes fitness for use (Juran 1980), conformance to requirements (Crosby 1979), and meeting the expectations of customer (Feigenbaum 1991). It appears that the traditional definition of quality is positive, i.e., measures customer satisfaction, while Taguchi uses the negative approach, i.e., losses to the customer. Note that in practice product negatives are more actionable than "product positives."

In this section, we introduce the Taguchi method as another tool for optimization. It is not intended to be a complete presentation of the method, but rather to show that a hybrid between the traditional approach to optimization and the Taguchi method can be a valuable tool for R&D and manufacturing.

The Taguchi Method

The Taguchi approach to obtain robustness of products and processes is a sound concept for achieving quality. The central concept of the Taguchi method is based on the following:

1. Robustness of products, services, and processes. This implies that a product should perform consistently in various conditions of use by the consumer.
2. Use of experimental design that incorporates both the assignable (controllable) and unassignable (uncontrollable) causes of variation. Taguchi contends that to accomplish robustness, these two causes of variation should be included in the experimental design.
3. Use of target value instead of the traditional process specifications. This concept has a strong bearing in manufacturing processes. Taguchi contends that it is difficult to obtain process variability, i.e., standard deviation, during normal production, but one can obtain variability as a deviation from the target value. Furthermore, the deviation from target value can be translated into a society/consumer monetary loss, the measure of which is given by the loss function statistic. It is this measure that the Taguchi method was able to link engineering results to management thinking, i.e., profit and loss.

Let us discuss the techniques that Taguchi recommends to achieve the central basis of his method. The focus of this section will be on R&D formulation problems and the use of sensory data to define perceived quality.

Types of Quality Characteristics

Quality characteristic in the Taguchi method can be viewed as a response variable as discussed in Section 8.1. Taguchi (Taguchi and Wu 1979; Taguchi 1986) classified the response variable into five types, but only three of these will be given here because of their applicability to R&D formulation work and in sensory testing. In sensory testing, the term perceived quality is used to denote the subjectivity of the measurement process. The three types of quality characteristics are given below.

1. Nominal-Is-Best. This is a characteristic with specific quantitative target value. Examples are viscosity measurements in shampoo and conditioner, which should not be too low or too high; the level of fragrance in bar soaps; and the amount of salt in a food product.

2. Smaller-Is-Better. This is a characteristic where the target value of the response is zero or at its minimum. The amount of moisture loss on the skin is an example. For a skin moisturizer, the product should keep moisture losses from the skin to a minimum.

3. Larger-Is-Better. Here the desired characteristic has the largest value. Examples are percentage yield, length of shelf-life, acceptance/preference values, and others. Here, we want to maximize the value of the response.

A unique idea of the Taguchi method is that the quality characteristic should be measured as a function of the controllable factors rather than its symptoms (response). For example, in studying the clearance between the door and the body of a car, the clearance should be nominal to provide an easy opening of the door and at the same time prevent undue friction. Here, one measures the distance between the door and the body as a quality characteristic and the distance/clearance is a function of quality. One does not measure the friction, which is the symptom; rather one measures the cause of the symptom.

Problems with Perceived Quality Characteristics

It is known that sensory quality is primarily a function of the ingredients in the formula. In sensory evaluation and consumer testing, the symptoms of quality are the one being measured because of the poor physical correlation between function and consumer acceptance. This contradicts Taguchi's basic idea of quality measurement. Attempts to measure hedonic values by instrumentation have not been successful.

In the absence of physical correlates with sensory quality, consumer studies use humans as instruments to measure perceived quality characteristics. It is known that only the consumer can meaningfully measure hedonic values, and this is where instrumentation has failed. Figure 8.13-1 shows an approach to using perceived quality to fit into the Taguchi method. The establishment of parameter design of the products is done by consumer feedback. Once established, these parameters become the target value to monitor product quality, and at this stage one can go directly from point A to point D. The drawback of this approach is that consumer feedback is inherently subject to large variability. However, there is a wide tolerance of consumer product acceptance that may compensate for this variation. Note also that there is no wrong answer in hedonic responses, and thus one must inevitably face this variability.

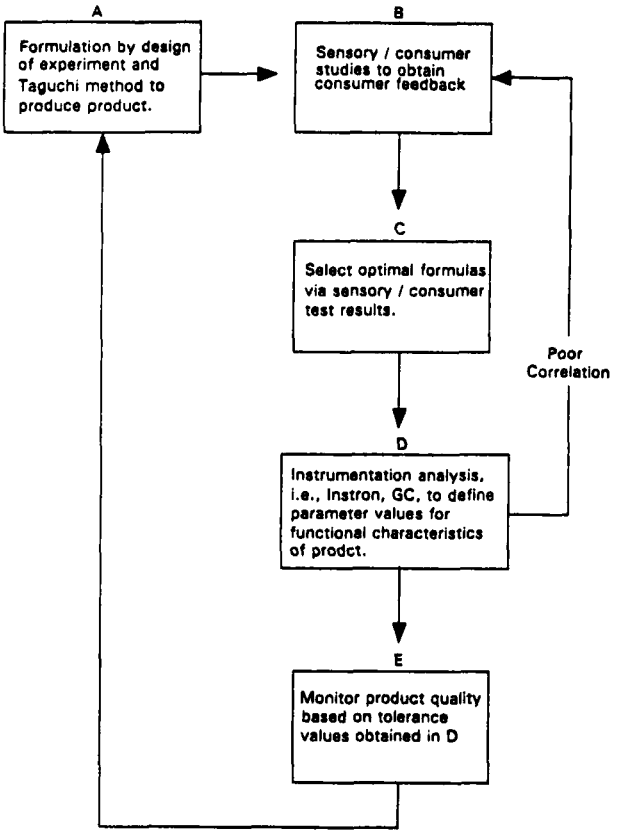


Fig. 8.13-1
Use of consumer feedback as a quality measure for the Taguchi method.

When the parameter design is completely specified, the variation in the amount of ingredients going into the product during manufacture should be minimized from the target value. This is where instrumentation analysis comes into the whole total quality process.

The Measurement of Quality

The quality of a product or process should be optimized in terms of cost and consumer acceptance. The initial step in quality work is to set the parameter design defined by Taguchi as the parameter values of a product or process so that the resulting product is functional with a high level of performance and is minimally sensitive to noise. In product formulation work, the parameter values are the amount of each ingredient in the product established by appropriate formulation design. The optimization of parameter values is evaluated by Taguchi's signal-to-noise ratio, denoted in this book by SNR, and by the loss function statistic denoted by L. The SNR provides the optimal parameter values, and the L statistic provides a measurement of quality as a loss to society. The SNR and L values are therefore directly related to the measurement of quality.

In the measurement of quality through parameter design, the controllable and the uncontrollable factors are built into the product. The controllable factors are those factors that can be controlled or varied by the experimenter, such as ingredients in the formula. The uncontrollable factors are the noise affecting the functional characteristics of the product, such as consumer abuse during product usage, water hardness, temperature, humidity, manufacturing imperfection, product deterioration/stability, and others; these are noise factors beyond the control of the manufacturer and to control these factors is obviously futile and not recommended. This is one example of the uniqueness of the Taguchi approach to total quality.

There are experimental designs that can be used to obtain an optimal parameter design of a product. They are as follows:

1. Taguchi's orthogonal array (Taguchi and Konishi 1987).
2. Full factorial and fractional factorial designs as discussed in Chapter 6.
3. Plackett-Burman design discussed in Section 8.4.
4. Central composite design discussed in Section 8.6.
5. Mixture design discussed in Section 8.7.

The orthogonal arrays, fractional factorial, and the Plackett-Burman designs are related in many ways, but such relationships are not discussed here. See Bullington *et al.* 1990; Kacker *et al.* 1990, 1991; Montgomery 1991). We briefly discuss the orthogonal arrays together with examples.

Orthogonal Arrays. Taguchi's orthogonal array design has a sound mathematical basis (Kacker *et al.* 1990, 1991) with similarities to fractional factorial designs. The

confusion between the orthogonal array and the traditional fractional factorial designs stems from equating the two designs to be the same, instead of being similar in nature.

A catalogue of orthogonal arrays is found in Taguchi and Konishi (1987), and the most commonly used of these arrays are shown in the appendix Table N. The convention for naming arrays is $L_a(b^c)$ where

a = number of experimental runs, i.e., number of formulations.
b = number of levels of each factor, i.e., low and high; low, medium, and high.
c = number of columns in the array, i.e., number of factors/ingredients to be studied.

The letter L derives from the Latin square design because of its mathematical connection to Latin square. See Gacula and Singh (1984), among others, for a discussion of the Latin square design.

Consider an $L_4(2^3)$ orthogonal array:

Formula	A	B	C
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1

To study the effects of three factors A, B, and C, one can run four test formulas where each factor is varied in a balanced arrangement, that is, they are orthogonal; each factor with two levels (1 = low, 2 = high) occurring equally often. The above array is called the inner array (IA) of the design. It can be shown that this array is similar to a 1/2 fraction of the 2^3 factorial design.

The uniqueness of the Taguchi design is the following:

1. Any column can be deleted without destroying the design.

2. The columns are merely a label, thus the factor can be assigned to any column to facilitate execution of the experiment. For example, if a factor is difficult to change or manipulate, this factor should be the first column in the array, followed by the second most difficult to change, and so on.

3. Inclusion of noise factors in the design. These factors are called outer array (OA). The inclusion of noise factors in the design results in products/processes to be robust and put less emphasis on randomization in the execution of the experiment. However, this book recommends the use of randomization as part of the experimental design. If a product is to be used by both males (M) and females (F), then sex is an example of a noise factor. Another example of a noise factor is water hardness. More than one noise factor can be included in a design; however, the

number of experimental runs becomes large. Remember that noise factors are those factors that cannot be controlled by the manufacturer when the product is at the hands of the consumer. In our example, the resulting design is as follows:

Formula	Inner array			Outer array	
no.	A	B	C	M	F
1	1	1	1	m1	f1
2	1	2	2	m2	f2
3	2	1	2	m3	f3
4	2	2	1	m4	f4

The outer array are observed values of the response for each sex. Note that with two noise factors, the total number of runs becomes eight, i.e., number of test formulas × number of noise factors. In this design, formula 1 has all factors at the low level, formula 2 has factor A at the low level and the remaining factors B and C at the high level, and so on. The SNR is computed for each formula across sex to determine the formula that is most insensitive (robust) to sex differences.

The Larger-Is-Better. The success of the Taguchi method leans heavily on the choice of quality characteristic to be measured. In larger-is-better, the target value m = ∞. The L function for one unit is given by

$$L(y) = A_o y_o^2 (1/y^2), \qquad y \geqslant 0 \tag{8.13-1}$$

and for more than one unit

$$L(y) = k(MSD)$$

where

$$k = (\text{consumer loss } A_o)(\text{consumer tolerance } y_o)^2$$
$$MSD = \text{Mean squared deviation}$$
$$= \Sigma(1/y_i^2)/n, \qquad i = 1, \ldots n \text{ units}$$

Note that the mean squared deviation is the inverse of the observation y resulting in a quadratic curve for the loss function. The relationship of the terms in Eq. (8.13-1) is shown in Fig. 8.13-2. Notice that as the value of y increases, the consumer loss A_o decreases. This is obvious because we are targeting a larger value (m) of the characteristic. Likewise, as the value of y decreases, the loss to the customer increases.

For large sample size n, MSD can be approximated by

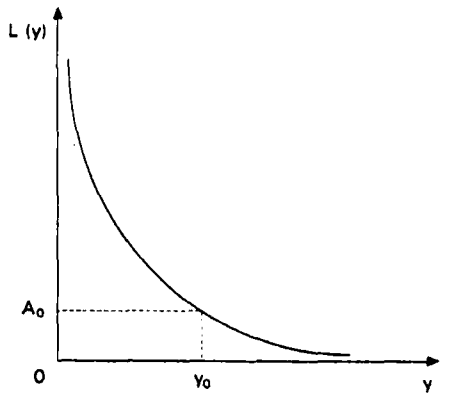


Fig. 8.13-2
Loss function for "larger-is-better."

$$MSD = (1/\bar{y}^2)[1 + (3s^2/\bar{y}^2)]$$

where $\bar{y}$ and s^2 are sample mean and sample variance, respectively. A better approximation of MSD is reported by Maghsoodloo (1990).

The signal-to-noise ratio, in decibel unit db, is obtained by the formula

$$SNR = -10\log(MSD) \tag{8.13-2}$$

which is the negative of the logarithm of MSD calculated for each run (treatment combination). In this ratio, we want to maximize the effect of the treatment (signal) and minimize the noise (random/systematic error), hence, large values of SNR are to be desired. Since the appropriate sample size n plays an important role in the accurate estimation of MSD, it is suggested that the traditional method for calculating sample size should be considered, i.e., see Gacula and Singh (1984).

Scales for Perceived Quality

The most popular scale of measurement to obtain consumer acceptance of products/services is the hedonic scale (Peryam and Girardot 1952). This scale defines the psychological states of like and dislike on a linear scale with like on the upper end and dislike on the lower end of the scale. Variations in the number of categories on the scale have been reported, but the 9-point scale is recommended in this book. See Gacula and Singh (1984), Stone and Sidel (1984), and Meilgaard *et al.* 1987) for further reading.

Table 8.13-1 shows the hedonic scale and the corresponding SNR for each category. Obviously, the desired target score is "larger-is-better." See Fig. 8.13-3 for the plot of scale y and SNR.

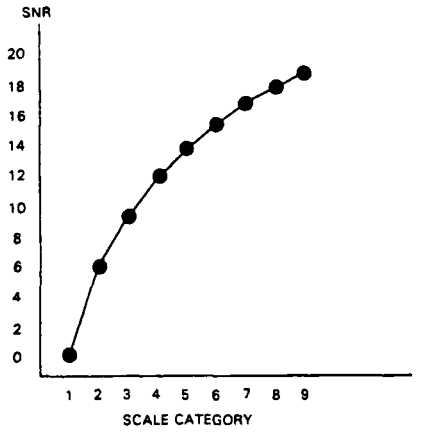


Fig. 8.13-3
The plot of scale value and SNR for the hedonic scale.

Table 8.13-1
The 9-point hedonic scale as used in the Taguchi method for larger-is-better quality characteristic.

Scale category, y	$1/y^2$	SNR
1: dislike extremely	1.0000	0
2: dislike very much	0.2500	6.021
3: dislike moderately	0.1111	9.543
4: dislike slightly	0.0625	12.041
5: neither like nor dislike	0.0400	13.979
6: like slightly	0.0278	15.560
7: like moderately	0.0204	16.904
8: like very much	0.0156	18.069
9: like extremely	0.0123	19.101

Note: $SNR = -10\log[(1/y^2)/n]$, where $n = 1$.

Another scale used to obtain consumer feedback for intensity of a characteristic is the so-called just right scale. In this book, the 5-point scale is recommended. Table 8.13-2 shows the scale along with its corresponding SNR and the plot shown in Fig. 8.13-4. In this scale the target value is category 3 ($m = 3$), hence belongs to Taguchi's "nominal-is-better" quality characteristic to be discussed later.

Example 8.13-1

Let us use the data of Example 6.1-1 (Chapter 6) with modification as shown in Table 8.13-3. In this table, A and B are the controllable factors and S1 and S2 are

Table 8.13-2
The 5-point just right scale as used in the Taguchi method for nominal-is-better quality characteristic.

Scale category, y	$(y - 3)^2 + k$	SNR
1: much too strong	4.0 ± 0.25	−6.284
2: too strong	$1.0 + 0.25$	−0.969
3: just right	$0.0 + 0.25$	6.021
4: too weak	$1.0 + 0.25$	−0.969
5: much too weak	$4.0 + 0.25$	−6.284

Note: $SNR = -\log[\Sigma(y - 3)^2/n]$, where $n = 1$. The constant k is used to avoid taking the logarithm of zero which is undefined. The value of k can range from 0.1 to 1.0.

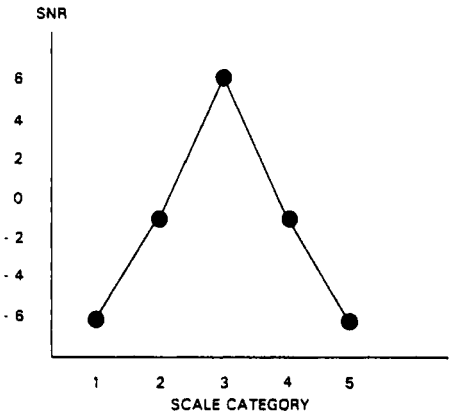


Fig. 8.13-4
The plot of scale value and SNR for the just-right scale.

the uncontrollable factors (noise). The noise factors are types of bacteria. When products are used by the consumer, manufacturers do not have control of the kind of bacteria that the consumer are exposed to, hence one must develop a robust product that can control as many as possible of the bacterial types. In this example, the quality characteristic is the percentage of bacterial reduction from placebo, the larger reduction the better.

As shown in Table 8.13-3, run #4 with an SNR = 31.08 db is the largest, thus this run, which consists of the high levels of both factors (A2B2), is the optimal formulation for the control of bacterial growth.

Although it is difficult to assign dollar value, let us do so for the sake of illustration. Suppose that for every 5% decrease in effectiveness results in $1.00 loss to

Table 8.13-3
Calculations for Example 8.13-1 to illustrate the larger-is-better quality characteristic.

	Controllable factors			Noise factors	
Run	A	B	AB	S1	S2
1	−1	−1	−1	13.0	8.0
2	1	−1	1	32.6	30.0
3	−1	1	1	22.4	18.5
4	1	1	−1	38.0	34.0
Sum of (−1)	61.9	83.3	93.0		
Sum of (+1)	134.6	112.9	103.5		
Contrast	72.7	29.6	10.5		

Calculation of mean squared deviation:
$MSD_1 = [1(1/13.0^2) + (1/8.0^2)]/2 = 0.01111$
$MSD_2 = 0.00102$
$MSD_3 = 0.00245$
$MSD_4 = 0.00078$
Calculation of signal-to-noise ratio:
$SNR_1 = -10\log(0.01111) = 19.55$
$SNR_2 = 29.91$
$SNR_3 = 26.11$
$SNR_4 = 31.08$

the customer, what is the loss function? Here, $A_o = \$1.00$ and $y_o = 5\%$, therefore for run #1 the loss is

$$L(y_1) = 1.00(5^2)(0.01111)$$
$$= \$0.28 \text{ per unit}$$

For the optimal run (#4),

$$L(y_4) = 1.00(5^2)(0.00078)$$
$$= \$0.02 \text{ per unit}$$

The result is clear: the consumer will lose only 2 cents per unit using the optimal formula, whereas loss is 28 cents for the worst formula. If a company sells 1,000,000 units a year, then the loss to society (customer) for the worst formula (#1) is

$$1{,}000{,}000(0.28) = \$280{,}000.$$

The design matrix of this example is L_4 on the Taguchi orthogonal array design:

L_4	B	A	AB
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1
Sum low:1	83.3	61.9	93.0
Sum high:2	112.9	134.6	103.5
Delta	29.6	72.7	10.5

In the Taguchi notation, the low level is denoted by 1 and the high level by a 2, and this corresponds to -1 and 1 on the factorial design notation, respectively. Notice that the absolute value of delta is called contrast in the factorial design terminology.

The Smaller-is-Better. In the smaller-is-better quality characteristic, the target value or the ideal value for y is m = 0, or the smallest value on a measurement scale. The loss function for one unit is

$$L(y) = (A_o/y_o^2)y^2, \qquad y \leqslant 0 \tag{8.13-4}$$

and for more than one unit

$$L(y) = (A_o/y_o^2)MSD$$

where

$$MSD = (y_1^2 + y_2^2 + .. + y_n^2)/n$$

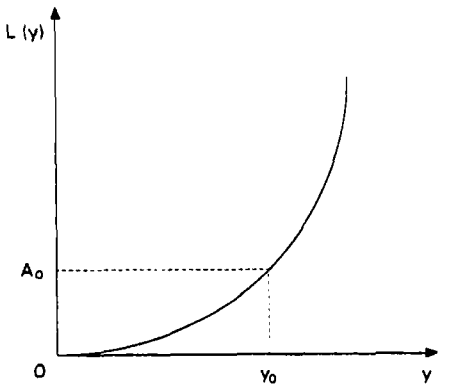


Fig. 8.13-5
Loss function for "smaller-is-better."

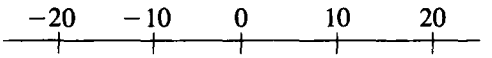
and for large n, MSD can be approximated by $MSD = \bar{y}^2 + s^2$

where $\bar{y}^2$ is the mean squared and s^5 the sample variance calculated by $\Sigma(y_i - \bar{y})^2/n$. Fig. 8.13-5 shows the loss function for this characteristic. Again, it is quadratic in y and is a mirror image of the larger-is-better. The closer the value of y to m, the lesser is the loss to the customer.

The signal-to-noise ratio is the same in form to that of the larger-is-better which is

$$SNR = -10\log(MSD)$$

In relation to the loss function, the SNR is a decreasing function of L(y); i.e., larger values of SNR will always lead to smaller values of loss per unit. Since y can be less than 1.0, the SNR values can be either positive or negative, and large value of SNR is again desirable. The linear scale representation is given below:



Thus, an SNR of −10 is better than an SNR of −20; an SNR of 20 would be the best on the above representation.

Example 8.13-2

This example deals with percentage weight loss of beef feeders. The cattle feeder desires to minimize weight loss of beef feeders during transit on its way to the buyer. The buyer deducts weight loss during transit in calculating the final weight of the animal for payment to the producer. Reducing the weight loss benefits both the producer and the buyer. There are four factors considered important:

A = Ratio of concentrate to roughage in the feed, 50/50 vs 70/30.
B = Days on feeding test, 120 vs 130 days.
C = Tranquilizers during transit, 10 mg vs 50 mg daily.
D = Type of rations, pellets vs meal.

The noise factor is the distance of travel, N_1 = short and N_2 = long.

Table 8.13-4 shows the percentage weight loss during transit of 16 beef feeders. In this table, design matrix for both the fractional factorial (1/2 of 2^4) and the orthogonal array (L_8) are shown. Notice that the columns for the array in this table correspond to columns 4, 2, 1, and 7 of L_8.

The largest SNR is $SNR_8 = 4.81$, which consists of the treatment combination $A_2B_2C_2D_2$. On this basis, the optimum combination consists of all factors at high level with an average weight loss during transit of 0.57%. The delta values indicate that factor D is the most important factor affecting weight loss. Based on the SNR

Table 8.13-4
Percentage weight loss of beef feeders to illustrate the smaller-the-better quality characteristic.

	Fractional factorial				Inner array				Outer array		
L_8	A	B	C	D	A	B	C	D	N_1	N_2	Sum
1	−	−	−	−	1	1	1	1	1.31	1.25	2.53
2	+	−	−	+	2	1	1	2	0.83	0.98	1.89
3	−	+	−	+	1	2	1	2	1.09	0.78	1.87
4	+	+	−	−	2	2	1	1	1.29	1.17	2.46
5	−	−	+	+	1	1	2	2	0.97	0.80	1.77
6	+	−	+	−	2	1	2	1	1.16	1.29	2.45
7	−	+	+	−	1	2	2	1	1.21	1.31	2.52
8	+	+	+	+	2	2	2	2	0.53	0.61	1.14
Sum of low level					8.69	8.64	8.75	9.96			
Sum of high level					7.94	7.99	7.88	6.67			
Delta					0.85	0.65	0.87	3.29			

Calculation of mean squared deviation:

$MSD_1 = [(1.31)^2 + (1.25)^2]/2 = 1.64$
$MSD_2 = 0.82$
$MSD_3 = 0.90$
$MSD_4 = 1.52$
$MSD_5 = 0.79$
$MSD_6 = 1.50$
$MSD_7 = 1.59$
$MSD_8 = 0.33$

Calculation of signal-to-noise ratio:

$SNR_1 = -10\log(1.64) = -2.15$
$SNR_2 = 0.86$
$SNR_3 = 0.46$
$SNR_4 = -1.82$
$SNR_5 = 1.02$
$SNR_6 = -1.76$
$SNR_7 = -2.01$
$SNR_8 = 4.81$

and the delta values, one may conclude that the optimum factor combination may be $A_1B_1C_2D_2$ and/or $A_2B_2C_2D_2$, and confirmation runs should be done on these combinations.

Assuming that for every weight loss of 1.0% results in a \$12.00 loss, the L function for the worst SNR is

$$L(y) = (12.0/1.0^2)1.64 = \$19.68$$

and for the best SNR is

$$L(y) = (12.0/1.0^2)0.33 = \$3.96.$$

It is clear that reducing the variability by optimizing the treatment combinations results in cost savings to both the producer and the buyer, and indirectly to the consumer through lower production costs.

The Nominal-Is-Best. The loss function L(y) for nominal-is-best (aim for target) is shown in Fig. 8.13-6. The target value m is the nominal value of y, i.e., m = y. The loss function is

$$L(y) = (A_o/y_o^2)(y - m)^2 \tag{8.13-5}$$

which generalizes to

$$L(y) = (A_o/y_o^2)(MSD)$$

for more than one unit, where MSD is given by

$$MSD = [(y_1 - m)^2 + (y_2 - m)^2 + .. + (y_i - m)^2]/n,$$
$$i = 1, 2, .., n$$

For large sample size

$$MSD = s_n^2 + (y - m)^2$$

An important result at this point is that the larger-is-better, the smaller-is-better, and the nominal-is-best measure both the variability (MSD) of the response and the closeness of the mean response (y) to the target value.

The signal-to-noise ratio is

$$SNR = -10\log(s_n^2 + (y - m)^2)$$
$$= -10\log(MSD)$$

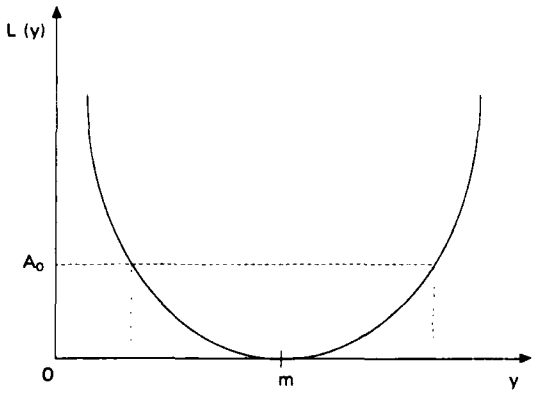


Fig. 8.13-6
Loss function for "nominal-is-best."

Example 8.13-3

Five factors denoted by A, B, C, D, and E are ingredients believed to enhance the flavor of a beverage product. A 5-factor Plackett-Burman design was used to evaluate the effects of these ingredients on sweetness. The 5-point just-right scale (1 = much too weak, 3 = just right, and 5 = much too sweet) was used by 10 expert judges to evaluate each sample for sweetness. In this scale the target value m = 3. Table 8.13-5 shows the design matrix and the response under two noise factors: N_1 = refrigerated vs N_2 = frozen; these factors are suitably a noise because the manufacturer intends to produce a product that will perform satisfactorily to either storage conditions.

The use of means on scales with the middle category as the optimal is not recommended, because variables can have a mean of equal to 3.0 (just right) but one of them may have a wider range than the other. This is clearly shown in this example. Based on the mean scores, run #5 hits the target value of 3.0; the observations for this run range from 2 to 4. Run #2 with a mean of 3.2, although not exactly on

Table 8.13-5
Beverage study to illustrate nominal-is-better and the just-right scale.

	Factor					Noise			
Run #	A	B	C	D	E	N_1	N_2	Total	Mean
1	1	1	1	−1	1	3 2 3 4 3	3 3 3 2 3	29	2.9
2	1	1	−1	1	−1	3 3 3 4 3	3 3 3 4 3	32	3.2
3	1	−1	1	−1	−1	3 3 3 3 3	2 3 4 3 2	29	2.9
4	−1	1	−1	−1	1	3 2 2 2 3	2 3 2 2 3	24	2.4
5	1	−1	−1	1	1	4 3 3 3 3	2 3 4 3 2	30	3.0
6	−1	−1	1	1	1	2 3 2 2 3	2 3 2 3 3	25	2.5
7	−1	1	1	1	−1	2 2 3 2 3	3 2 2 2 2	23	2.3
8	−1	−1	−1	−1	−1	3 2 3 2 2	2 2 2 2 3	23	2.3
Low	95	107	109	105	107				
High	120	131	106	110	108				
Delta	25	24	3	5	1				

Run #	MSD	SNR	
1	0.1340	8.7290	
2	0.1014	9.9396	(most robust)
3	0.1340	8.7290	
4	0.1944	7.1130	(least robust)
5	0.1290	8.8941	
6	0.1806	7.4328	
7	0.1524	8.1702	
8	0.1524	8.1702	

Note: Delta = High − Low, i.e., 120 − 95 = 25

target has a range of 3 to 4, obviously with lower variability than run #5. The lower variability for run #2 results in higher signal-to-noise ratio, therefore it is the optimal combination of the five factors.

Due to the aforementioned problem in using means, the frequency for the just-right category is the appropriate value to use to compare treatments/products. For example, the least robust run is #4 with SNR = 7.1130 and the percentage of scores on the just-right category is 4/10 = 40.0%. For run #2, it is 8/10 = 80.0%.

The Use of Signal-to-Noise Ratio in Formula Selection

The ultimate goal of testing various product formulas by the consumer is to obtain a high quality product under various conditions of use. That is, we want a robust product as measured by consumer feedback. For example, we want a product that scores high on most, if not all, sensory attributes. In this book, the Taguchi philosophy has been extended to sensory/consumer research. In this extension, the sensory characteristics are defined as "noise factors" as shown in Fig. 8.13-7. The unique-

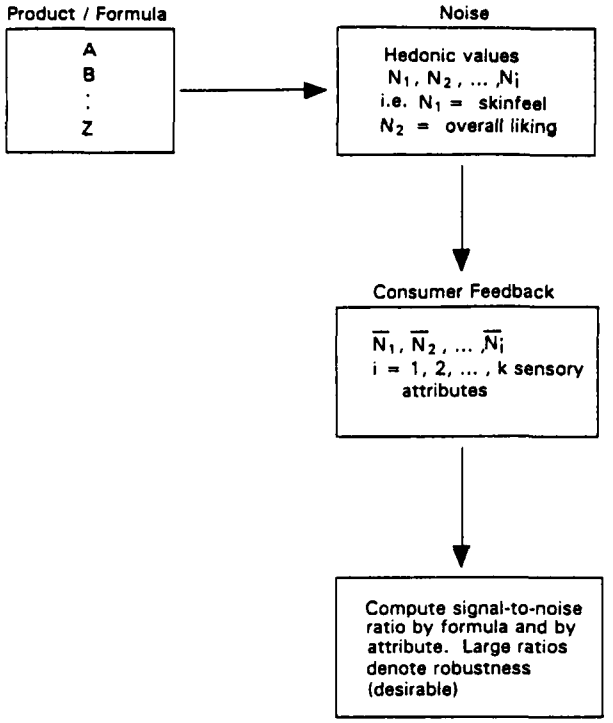


Fig. 8.13-7
Central concept of the extension of Taguchi's method to formula selection in consumer research.

ness of this extension is obtaining the SNR by product formula and by attribute. The SNR by formula provides measurement of product robustness against sensory attributes, whereas the SNR by attribute provides consistency of individual consumer responses. In both cases, we want to maximize the value of SNR.

Figure 8.13-8 depicts what is meant by a robust product. The axial arm represents an attribute, and the length of the arm represents the mean score measured from the origin with value of 1 and a maximum of 9 for the 9-point hedonic scale. A perfectly robust product always forms a circle, i.e., all attributes are at the maximum. However, this is unattainable in practice. Plotting mean scores inside the circle forms the so-called cobweb diagram (Stone *et al.* 1974), which has been very useful in the presentation of results. Cliff and Wild (1990) have discussed SAS programs to generate these diagrams. It can also be drawn by Microsoft Draw (Microsoft Corp. 1992) as shown in Fig. 8.13-8. It is obvious in this figure that if a product is deficient in one or more attributes, a dented circle will result.

The traditional approach to product selection has been the use of significance testing. An improved approach is to combine significance testing and Total Quality methods of Genichi Taguchi. Products may not be significantly different, but they may differ in quality due to one of them having lower variability. To provide a tool

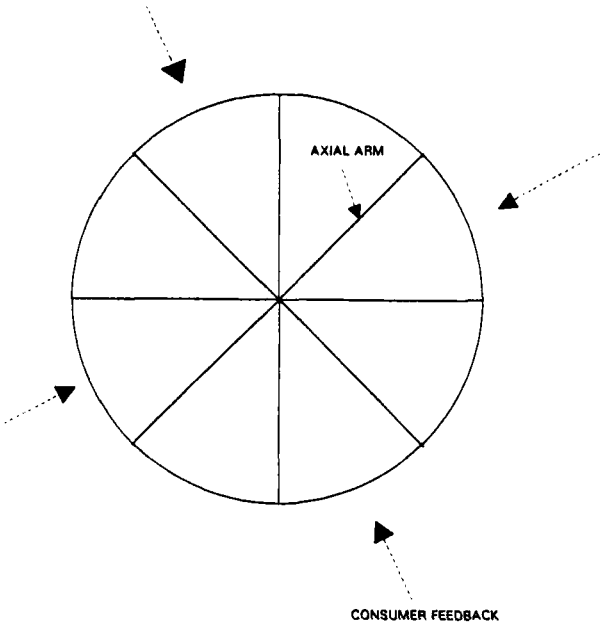


Fig. 8.13-8
Depiction of a robust product with consumer feedback defind as "noise."

for deciding which product should be selected, the SNR can be used. As applied to sensory research, the SNR can results in the following cases:

Case 1: Low mean and low SNR.
Case 2: Low mean and high SNR.
Case 3: High mean and low SNR.
Case 4: High mean and high SNR.

Case 1 is obviously undesirable; it did not meet the total quality goals (high mean and high SNR). Case 2 indicates low variation in the data meeting the goal for high SNR, but did not meet the other goal on high means (larger-is-better). In this situation, product reformulation is needed. However, if this was a process manufacturing study, high SNR should be the basis for decision and later on improve on the process mean. Case 3 indicates the presence of high variation in the data, which is likely due to consumer segmentation. When this case occurs, the decision should be based solely on mean score or analyze the data by segments/subgroups; one should examine the cause of segmentation. Case 4 is the ultimate goal. It meets both goals for "larger-is-better" perceived quality characteristic and low variation (high SNR). In consumer testing, a high mean generally denotes a high SNR in the absence of segmentation.

In looking at the SNR by attributes, one must examine the frequency distribution to facilitate proper interpretation of the result. The SNR statistic is extremely sensitive to values very distant from the target. For example, a hedonic distribution with one rating falling in the "dislike extremely" category of the scale will have

Table 8.13-6
Frequency distribution for overall liking to illustrate the sensitivity of SNR to observations far from the target (larger-is-better).

Scale category	Product A	Product B
1: dislike extremely	1	0
2: dislike very much	0	0
3: dislike moderately	4	5
4: dislike slightly	4	8
5: neither like nor dislike	19	22
6: like slightly	18	22
7: like moderately	31	32
8: like very much	40	25
9: like extremely	12	17
Mean score	6.80	6.62
SNR	14.71	15.37

Note: Mean scores for overall liking are not significantly different from each other.

a large impact on lowering the value of the SNR (Table 8.13-6). If this rating is, in fact, an outlier it should be excluded from the computation. When this situation occurs, the SNR by averages and the SNR by attributes often do not agree. In Table 8.13-6, although the mean score for Product A was higher than that of Product B, the SNR is slightly lower because of that one rating falling in the extreme category. When this rating is excluded from the computation, the SNR changed from 14.71 to 15.81 indicating that Product A is now a better quality than Product B. The presence of outlier in the data, especially those outliers well beyond the target, can severely distort the reliability of SNR. In particular, when the rank order of means and SNR is in disagreement one must examine the frequency distribution to search for outlying observations or possible segmentation, i.e., multi-modal distribution.

Let us consider the data in Table 8.13-7. Notice that three panelists rated Product A in the dislike extremely category, while none rated Product B in that category. Also, their mean scores are the same. As a decision maker, which product would you choose? For the hedonic scale, the quality characteristic is "larger-is-better," and the scores in the dislike extremely category is obviously off target. The SNR is found to be 12.9 for Product A and 15.2 for Product B, a result which clearly indicates that Product B is perceived to be of better quality. As a decision maker, one should choose Product B, even though the product means were not statistically different from each other.

Assume that the frequency distribution for Product A can be projected to a larger population, i.e., there will be 3 consumers for every 100 that dislike the product extremely. For 100,000 consumers, this translates to 1,000 of them disliking Product A. This is evidently the penalty for not meeting customers' expectations.

In another case, it is possible to have a product with slightly higher mean, but of lower quality compared to a product with lower mean. This is shown in Table

Table 8.13-7
Frequency distribution for fragrance liking to illustrate the use of SNR for quality decision.

Scale category	Product A	Product B
1 dislike extremely	3	0
2 dislike very much	1	3
3 dislike moderately	5	2
4 dislike slightly	10	11
5 neither like nor dislike	9	7
6 like slightly	16	21
7 like moderately	26	35
8 like very much	44	33
9 like extremely	16	18
Mean scores	6.8	6.8
SNR	12.9	15.2

8.13-8 for Products B and C. Products A and B are at parity (not statistically different at the 5% level), but A is of better quality; similarly, Products B and C are at parity, but C a better quality. Again, as a decision maker Product C would be chosen over B on the basis of quality.

Table 8.13-8
Frequency distribution for softness of skin after application of products A, B, and C.

Scale category	A	B	C
1		1	
2		1	1
3		3	
4	7	5	9
5	8	5	11
6	7	12	13
7	23	32	19
8	41	32	40
9	14	11	5
Mean scores	7.3a	6.9ab	6.8b
SNR	16.51	14.38	15.62

Note: Means with one letter in common are not significantly different from each other by the Bonferroni multiple comparison test (5% level).

Table 8.13-9
Hedonic mean scores for 10 sensory attributes and signal-to-noise ratio by products.

	Product			
Attribute	1	2	3	4
1. Fragrance before use	7.1a	7.0ab	6.9ab	6.7b
2. Fragrance during use	7.0	7.0	6.8	6.8
3. Clean feeling after use	7.2	7.2	7.1	7.1
4. Lather	7.5	7.6	7.2	7.3
5. Rinsing clean	7.2	7.1	7.2	7.1
6. Not drying skin	6.9	6.6	6.7	6.9
7. Moisturizing skin	6.6	6.3	6.4	6.7
8. Scent of skin	7.0	7.1	6.8	6.9
9. Feel of skin	6.7	6.6	6.7	6.8
10. Overall liking	7.0	6.8	6.7	6.8
Signal-to-noise ratio	16.92	16.80	16.71	16.79
Rank	1	2	4	3

Note: Means with one letter in common are not significantly different at the 10% level by the Bonferroni multiple comparison test.

Table 8.13-10
Signal-to-noise ratio by attribute for Example 8.13-4.

	Product			
Attribute	1	2	3	4
1. Fragrance before use	16.21*	15.56*	14.14	14.79
2. Fragrance during use	15.60	16.16*	15.87*	15.34
3. Clean feeling after use	16.28	16.64*	15.96	16.32*
4. Lather	16.46*	17.19*	14.60	16.15
5. Rinsing clean	14.55	14.90	15.80*	16.33*
6. Not drying skin	13.39*	11.80	12.77	15.49*
7. Moisturizing skin	13.04*	12.10	12.53	14.29*
8. Scent of skin	15.81*	15.03	13.48	15.99*
9. Feel of skin	13.81	14.03*	12.81	14.71*
10. Overall liking	15.36*	15.10	13.28	15.60*
No. of times on top 2* / total	6/10	5/10	2/10	7/10

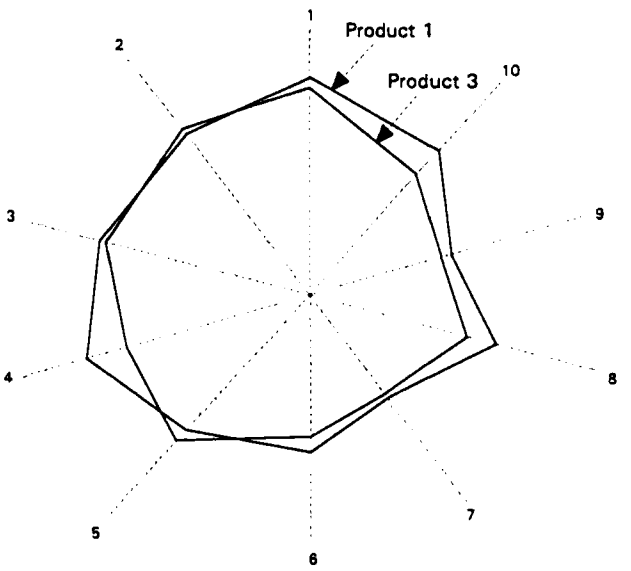


Fig. 8.13-9
Cobweb diagram for Products 1 and 3 to illustrate robustness.

Example 8.13-4

This example illustrates the use of signal-to-noise ratio in product/formula selection. Table 8.13-9 shows the mean scores for 10 sensory attributes obtained from a consumer test (N = 110) using the 9-point hedonic scale. Of the 10 attributes, only in "fragrance before use" were the four products significantly differentiated, Product 1 being more well-liked than Product 4. Numerically, the results tend to favor Product 1. Since the significance testing indicates an inconclusive result for product selection, the SNR can provide further information about the quality of each product as a basis for selection.

Using Eq. (8.13-2), the SNR based on means was computed and the results given at the bottom of this table. Although the SNRs were close, one can conclude that Product 1 with higher SNR should be selected. That is, it meets Case 4 of high means and high SNR (low variability). Products 2 and 4 are close alternative selections. In fact, Product 4 showed low variability (higher SNR) in consumer responses by attribute (Table 8.13-10). Product 3 should not be considered.

Fig. 8.13-9 shows the cobweb diagram for Products 1 and 3 to illustrate the striking differences between them. Although the diagram is dented, it shows that Product 1 is relatively robust.

CLAIM SUBSTANTIATION

We have discussed various experimental designs and their analyses useful in sensory testing and other research areas. In Chapter 8, experimental procedures were presented for obtaining an optimal formulation of products. Now we will compare an optimal product formula against a competitor or against a product class, the result of which will be used for advertising purposes. When we collect data for this purpose we are conducting a claim substantiation study. Results from this study can also be used to challenge competitors. Briefly, claim substantiation is defined as a statement of facts supported by evidence. This evidence is usually quantitative in nature and capable of being replicated under specified conditions. It is therefore evident that statistics and experimental design play a major role in furnishing evidence for establishing claims.

In this chapter, we present the statistical design and analysis to support claims substantiation. Although the design and analysis have been given in the previous chapters, our aim in Chapter 8 is to put these techniques in another framework for claims support.

9.1 CLAIM SUBSTANTIATION GUIDELINES

When one submits a claim for a product to television networks and newspapers, among others, it can be challenged by competitors through the National Advertising Division (NAD) of the Council of Better Business Bureaus (CBBB), a self-regulatory system for the industry. The evaluation of claims is based on its own review and evaluation of advertiser's substantiation and, when necessary, on consultation with technical experts. More importantly, an advertising claim is subject to government agencies standards, such as the Federal Trade Commission (FTC) and state governments, each of which may enforce differing standards. For example, as reviewed by Zelek (1990), FTC has stated that six factors should be taken into account in determining whether an advertiser has a reasonable basis of substantiation for advertising claims:

1. The product involved. Clearly, certain products require higher levels of substantiation. These include food, drugs and potentially hazardous products.
2. The type of claim. Where health or dietary claims are concerned, the level of substantiation for a reasonable basis will increase.

3. The benefits of a truthful claim. Where the benefits of a true claim are potentially great, less substantiation will be necessary.

4. The consequences of a false claim. As might be expected, where the potential consequences of a false claim are great, a higher level of substantiation will be required.

5. The ease of developing the necessary substantiation. When data required for substantiation is technically difficult or costly to generate, less will be required.

6. The amount of substantiation experts in the field would think is reasonable. As a rule, the substantiation must meet contemporary professional standards for evaluating the claim.

Claim substantiation cases and their resolutions are published in NAD Case Report, a CBBB publication. This publication is highly recommended reading material for sensory professionals involved in claim substantiation. An example of a substantiation challenge involving house paint (CBBB, Inc. 1990) is given in Fig. 9.1-1. The role of statistics, sensory, and experimental design is clearly evident in this challenge.

The results of exploratory research work should not be used for claims. Oftentimes, when the result of exploratory research is promising, it is being used for convenience to support a claim. This use of the result is generally unacceptable. To avoid litigation and post-claim substantiation costs, exploratory results should be repeated for the sole purpose of substantiation.

For example, when an exploratory experiment results in an optimal formula, a confirmatory laboratory experiment should be conducted using the paired comparison design for the purpose of claim substantiation. This comparison may include the optimal formula 1 vs. competitor 1, optimal formula 1 vs. competitor 2, and so on. If instrumentation data are part of the claim, these data should also be collected using the paired comparison design. However, in order to make a strong case for the laboratory data, i.e., trained panel/expert data, it should be supported by actual use of the product by the consumer. The consumer test methods discussed in Chapter 7 can be used for this purpose.

Another common exploratory research is through a focus group. Results from a consumer focus group is not acceptable for claims because the information obtained is qualitative. The result should be confirmed by a quantitative research, i.e., CLT, RGT (Research Guidance Test).

In providing general guidance on the definition of good consumer research, the courts and other authorities have embraced the views of the research community, which are summarized below (Zelek 1990). In fact, this guideline applies to all types of research activities.

1. A properly defined universe. The target consumer should be clearly identified.

2. A representative sample of that universe. Accepted sampling procedure should be followed.

3. Clear, precise and nonleading questioning. This specifically applies to survey research.

PPG INDUSTRIES, INC.
Lucite House & Trim Paint
Young & Rubicam/San Francisco

Basis of Inquiry: A magazine advertisement claimed: "In an independent test, monitoring 350 homes across the nation for 24 years, Lucite House & Trim paint consistently resisted cracking and peeling longer than comparable paints from Glidden, Dutch Boy, Sherwin Williams and Benjamin Moore." The advertisement featured a photograph of an exterior wall with one side showing paint failure and labeled "Glidden," and the other side showing no failure and labeled "Lucite House & Trim Paint." The advertising was brought to NAD's attention by the Glidden Company. The challenger stated its belief that the testing featured in the advertisement may not reflect current product performance and that the photograph represented an atypical result. NAD initiated an inquiry under *NAD/NARB Procedures*, effective April 1, 1990.

Decision: The advertiser provided reports of an ongoing paint performance program which compares its paint to competitive products. The program was commenced in 1966 and has continued through changes in ownership of the Lucite brand. Periodically a number of homes are painted by independent contractors, in regions with varying climates, using commercially available samples of the advertiser's paint and a competitive paint on adjacent sections of each home. Independent expert evaluators annually rate the performance of each section for cracking, peeling and flaking using a modified ASTM scale. Results indicated a significant performance advantage for the advertiser's brand compared to the challenger's brand and other competitive brands in each attribute. Comparisons of the most recently painted homes confirmed the earlier results. The advertiser stated that the photograph in the advertisement accurately depicted product performance on an actual test home.

The challenger queried several aspects of the design and results of the test study including the selection of houses with chronic problems, possible bias by the evaluators, the modified rating scale, and the fact that the photograph exaggerated the differences in the paints estimated by the modified ASTM scale.

In its review, NAD recognized that the design was appropriate for a stress test of durability of the brands in real-life situations and that the consistency of the results throughout the test period provided confidence that the observed differences were meaningful.

NAD agreed the claims were substantiated. (#2831JC, closed 11/21/90)

Fig. 9.1-1
An example of a substantiation challenge.

4. Sound interview procedures followed by competent interviewers with no knowledge of the purpose of the study. This also applies to survey research.
5. Accurate reporting of the results.
6. Appropriate statistical analysis of the data.
7. The objectivity of the entire process.

9.2 TESTING OF CLAIMS HYPOTHESIS

The conventional approach for testing hypothesis in scientific work has been to set the null hypothesis equal to zero. For example, when studying the effects of a certain flavoring additive on sensory characteristics of products one may use the symbols μ_1 and μ_2 to denote, respectively, the mean of scores without and with the flavoring additive to formulate a null hypothesis

$$H_0: \mu_1 - \mu_2 = 0.$$

The H_0 states that there is no effect of flavoring additive. An alternative hypothesis is

$$H_a: \mu_1 - \mu_2 \neq 0.$$

which states that the difference between mean scores is not equal to zero, i.e., could be positive or negative. Based on the sign of the mean difference, one may either conclude superiority or inferiority of one treatment over the other. A process for reaching a decision about the validity of H_0 or H_a is called testing of hypothesis. In reaching a decision, one may commit the so-called Type I and Type II errors due to sampling variations as discussed in Section 1.3 of Chapter 1. In the conventional approach, it is not appropriate to prove the null hypothesis, rather one disproves it. To prove a null hypothesis is irrelevant as one does not know with certainty the value of the null hypothesis; it is a claim that the investigator suspects is false. Hence, on the basis of statistical evidence the null hypothesis is proved either false or has insufficient evidence to disprove it. In his 1935 classic book, *The Design of Experiments*, Fisher wrote the following:

> In relation to any experiments we may speak of this hypothesis as the null hypothesis and it should be noted that the null hypothesis is never proved or established, but is possibly disproved, in the course of experimentation. Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis.

If we fail to reject the null hypothesis at a specified significance level, one concludes that there is no real difference between treatment means or that the observations possibly came from the same population. Note that one cannot conclude that the mean values are equivalent. In practice, failure to reject the null hypothesis is generally given less attention and at times the results are shelved, particularly when it is desirable to show a difference. On the other hand, the rejection of the null hypothesis attracts further inquiry and a search for a plausible explanation of the outcome is pursued that may lead to further experimentation. This is the general course of action encountered in many research studies.

In other experimental situations, the "acceptance" of the null hypothesis is the desired result. This situation leads to a logical conflict with the traditional approach

of disproving the null. Furthermore, the experimental design can be faulty, resulting in erroneous acceptance of the hypothesis. Therefore, the formulation of the null hypothesis and its analysis must be revised to accommodate the desired result. Such revision has been the subject of several investigations in clinical trials (Westlake 1972, 1976; Metzler 1974; Dunnett and Gent 1977; Blackwelder 1982). These revisions are as follows:

1. Use of the confidence interval method instead of test of significance.
2. Formulation of the null hypothesis with a specified difference.
3. Use of power of the test.
4. Use of a control chart that sets limits based on a specified number of standard deviations, i.e., 3 sigmas.

These revisions provide procedures to support claims for parity, superiority or inferiority. These procedures are discussed and illustrated by examples in the succeeding sections.

9.3 EXPERIMENTAL DESIGN AND CLAIM SUPPORT

The key to successful claim substantiation is the use of a correct experimental design to support the claim hypothesis. As a review, an experimental design is a structured plan conceived before the experiment is to begin; this plan includes clear statement of purpose of the study, how treatments/interventions are to be applied, the number of experimental units to be used, i.e., number of panelists/judges, accountability in the conduct of the experiment, the unit of measurement to be used to generate the data, development of the questionnaire, and the determination of the population to be sampled (Gacula 1987). All aspects of this plan are important and must be thoroughly understood by the members of the project team.

An example of a claim is: "You can't beat the taste of diet Sprite" (NAD Case #2501). This is a superiority claim and must be compared to all products in its class or category. The sensory attributes to define taste in the claim must be fully stated in the design. To maintain the same value of Type I error in all product comparisons, the paired comparison design should be used as opposed to comparing all products in a randomized complete block design. The reason for this is that in a series of independent paired comparisons, the confidence level of each comparison is always 1.0 — Type I error, thus all product comparisons have obviously the same confidence level. In this claim, all product comparisons must be statistically significant at the α level of significance to support the superiority claim.

Another example is the implied superiority claim in the context of Wyckham (1987), such as "No other cereal you can buy has more natural food fiber than Kellogg's All-Bran." This claim entails very extensive chemical testing of all breakfast cereals in the market. The design consists of obtaining random samples of each brand for chemical analysis. To support this claim, the differences in fiber content between

Kellogg's and the other brands must be statistically significant at the α level by the independent t-test or some other test statistics.

The easiest to substantiate are claims that do not involve the competitors directly, such as "Brand X fights cavities effectively or Brand X is an antibacterial agent." To support this claim, one must only show that the product is significantly effective at the α level when compared to a placebo. A simple group comparison design is used for this type of a claim and the data analyzed by an independent t-test or other appropriate nonparametric test.

On the other hand, a parity or similarity claim such as "NutraSweet^{TM} tastes just like sugar" (NAD Case #2490) involves comprehensive chemical testings and sensory evaluations. Although no product brands are compared, one has to provide data that show the active ingredient of NutraSweet^{TM} to have a close identity with natural sugars on both taste and biochemical aspects. Note that the desired result of this claim is the acceptance of the null hypothesis. The statistical procedure to support this claim is given in the next Section 9.4.

9.4 TEST FOR EQUIVALENCE AND SUPERIORITY

For two treatments to be compared under the conventional null hypothesis, the Student's t statistic is usually used as the test statistic for the rejection or acceptance of the null. Rejection of the null hypothesis does not indicate the practicality of the size of difference between treatments, rather it only indicates that the difference is not equal to zero or there is evidence of the presence of real effects.

In other situations, such as in clinical bioequivalence or bioavailability studies the interest is in the acceptance of the null hypothesis, i.e., the new drug is as effective as the standard drug. The use of the conventional hypothesis testing in this situation leads to difficulties of interpretation because the desired result is the acceptance of the null, hence the popular test of significance procedure no longer applies. In this respect, it has been suggested that the confidence interval test is more appropriate (Westlake 1972; Metzler 1974; Shirley 1976; Blackwelder 1982). As stated by Cochran (1983), a confidence interval relates to the question "How large is the difference?" In this book, procedures used to validate the acceptance of the null will be called a test for equivalence or parity.

Associated with the confidence interval of a parameter is the degree of confidence that the length of the interval will include the true value of the parameter. If the interest is in the difference between mean values of two populations, and if the hypothesized difference lies within this interval, then the mean values are said to be equivalent at a confidence level of $(1 - \alpha)100\%$ where α is the level of significance test determined in advance of the experiment. Suppose we test the hypothesis, H_o: $\mu_1 - \mu_2 = \mu_D$. The confidence interval for the difference μ_D between population means μ_1 and μ_2 is known to be

$$\bar{d} - t_{\alpha/2,n-1}(SE) \leqslant \mu_D \leqslant \bar{d} + t_{1-\alpha/2,n-1}(SE)$$

or simply

$$\bar{d} \pm t_{\alpha/2}(SE) \tag{9.4-1}$$

where $\bar{d}$ is the estimate of μ_D, the difference between sample means from the two populations, t is a value obtained from the t distribution at α level of significance and appropriate degrees of freedom, and SE is the standard error of $\bar{d}$ obtained by $S/\sqrt{N}$, where S is the standard deviation of $\bar{d}$. For large N, the t distribution approaches the standard normal distribution of Z with mean zero and variance 1; for example, for $\alpha = 0.05$, the value of $Z = 1.960$ (Appendix Table L). Thus Eq. (9.4-1) can be written as $\bar{d}$ $Z_{\alpha/2}(SE)$. In order for the confidence interval method to be valid, the samples must be taken independently from each population. Moreover, the samples must be sufficiently large to compensate for the lack of normality and other distributional assumptions. The decision rule for test of equivalence is stated below:

> *If the interval of μ_D includes zero, one concludes that the two treatment means are equivalent with a confidence level of $(1 - \alpha)100\%$. If in particular $\mu_D = 0$, then the means are considered nearly equal.*

Although the interval includes μ_D, there are other important factors that should be considered for supporting an equivalency position. The first factor is the power of the test, defined in several books (Dixon and Massey 1957; Quenouille 1965; Gacula and Singh 1984; Lehmann 1986) as the probability of rejecting the null hypothesis when it is false, or conversely, the probability of rejecting the null hypothesis when the alternative is true. Since in a parity position one desires the acceptance of the null, one needs a powerful test to reject the null hypothesis when it is false. Probabilities associated with the power of the test range from 0.0 to 1.0, with higher values indicating a powerful test. The power of a test can be calculated at any point in the alternative hypothesis and can be plotted for ease of examination. The test with the acceptable power is used to support equivalence or parity. Since power is the probability of rejecting the null hypothesis when it is wrong, the more the power of a test, the better it is.

The second factor is the experimental design that specifies the sample size, randomization, replication, control of variation, and experimental execution. With a good experimental design, one can feel reasonably certain that a statistically nonsignificant result is evidence of parity.

The use of a plot similar to a quality control chart can also be used to support parity by plotting individual replication of the experiment. See Schilling (1982) and Ryan (1989) for the construction of a quality control chart. Three standard deviations are recommended as the parity limits because we assumed under the null

hypothesis that we are dealing with one population. Note that a three-standard deviation limit includes about 99% of observations in the population. If majority of the differences fall within the parity limits, then there is evidence of equivalence. The basic rationale of the application of control chart is sampling. If each replication represents a random sample from the same population with similar mean and variance, then most of the differences should fall within the specified parity limits. Figure 9.4-1 shows a control chart for substantiating a parity position. Equation (9.4-1) can be used to obtain the width of the parity limits, such that a 95% confidence limit is $\bar{d} \pm 2.0(SE)$ and a 99% confidence limit is $\bar{d} \pm 3.0(SE)$. We call the chart in Fig. 9.4-1 a $\bar{d}$ control chart to indicate that mean differences are plotted.

The decision rule for superiority claim can be stated as follows:

> *If the lower limit of the confidence interval of the difference is equal to or greater than zero, the treatment or product with the larger mean is declared superior at the* $(1 - \alpha)100\%$ *level of confidence.*

Note that the result of the confidence interval analysis can also be interpreted as a test of significance at the α level (Natrella 1960; Barr 1969; Jones and Karson 1972; Cochran 1983). In particular, Gacula and Singh (1984) applied the confidence interval significance test in paired comparison of scale values obtained by the Thurstone-Mosteller model. The interpretation of the result is the same as in the conventional hypothesis testing.

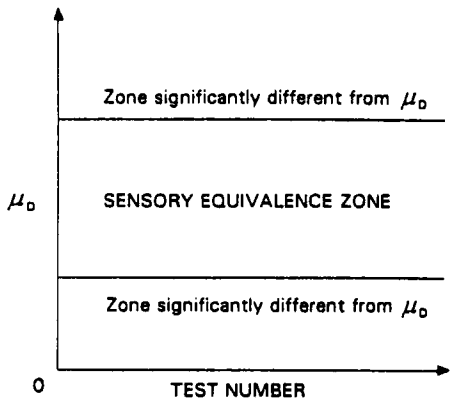


Fig. 9.4-1
A chart to support the position of sensory equivalence (parity) or superiority/inferiority.

Calculation of Power of the Test

As stated earlier, although a nonsignificant result may indicate parity, we need to know the power of the test, for it indicates the probability of rejecting the null hypothesis when it is false. In substantiating a parity position, it is desirable that this probability should be high for the alternative hypothesis. Unfortunately, the power of the test cannot be adequately estimated before the study because of the necessity of knowing the estimate of variance in the data, and this variance comes only from the study itself. As stated earlier, power ranges from 0.0 (no power) to 1.0 (high power). A power of 0.50 may be a reasonable value as the lower limit on the following basis. If we plot the power function on the vertical axis and the critical value $[Z_{\alpha/2}(SE)]$ on the horizontal axis, the power at the critical value is 0.50 regardless of α level and sample size. The choice of the size of power also depends on the severity of the consequence of an erroneous conclusion. In general, one must aim at higher power so that the study can withstand scientific and legal challenges.

In clinical trials involving drug safety and efficacy, a power of 0.90 has been reported (Frieman *et al.* 1978). Because of the high risk involved in clinical trials for making a wrong decision, the sample size of the study is determined in advance of the experiment by calculating the sample size based on the variance of historical data. The aim in this calculation is to obtain a power close to 1.0.

It is known in the statistical literature that three factors can increase the power of the test. These factors are as follows:

1. Type I error α. A test with large Type I error, i.e., 0.10, 0.20, will have more power. Thus, it is a trade-off between the significance level α and the power. For a fixed sample size test, it is not possible to have a high significance level and expect to increase the power. The relationship between the power of the test and significance level α or Type I error can be seen below:

Type I error	Confidence level	Power of test
0.01	0.99	Low
0.05	0.95	↓
0.10	0.90	
0.20	0.80	High

From the information above, it is evident that a compromise should be made between significance level and power.

2. Sample size. In sensory and consumer testing, sample size refers to the number of panelists or judges, computed to provide an adequate protection from the risks of α and β errors. The larger the sample size, the smaller the Type I (α) and Type II (β) risks. Once the compromise between the significance level and the power is made, the power of the test can be further increased with larger sample size.

3. Setting a reasonable difference between the value of the null and the alternative hypotheses. This factor is simply formulating a hypothesis with a specified difference. This factor is especially important for testing parity. In the absence of a method for specifying the difference, one resorts to determination of the difference based on historical data. For example, one may use the estimate of random error as the specified difference, or one can also use a fraction of the standard error of the mean or difference.

The power of the test can be computed by

$$\text{Power} = 1 - \beta \tag{9.4-2}$$

where β is the probability (P) of committing the Type II error. Assuming that we are sampling from a normal distribution and desire to test

$$H_o: \mu_1 - \mu_2 \leqslant \mu_o$$

against

$$H_a: \mu_1 - \mu_2 > \mu_o$$

then

$$\beta = \text{P(accepting } H_o \text{ when } \mu_1 - \mu_2 \neq \mu_o).$$

Here, β is the probability of incorrectly accepting the null hypothesis when in fact the alternative is true. An understanding of both Type I and II errors can be obtained from graphical illustration in Fig. 9.4-2. The first step is to find the critical value Z_c of the test statistic $Z = (\bar{X}_1 - \bar{X}_2 - \mu_o)/SE$ that divides the distribution into the so-called acceptance and rejection regions (Fig. 9.4-2a). If the value of Z lies in the rejection region, we reject the null hypothesis (H_o); conversely, if it lies in the acceptance region we accept it.

The rejection region is specified once the level of significance α of the test is known, usually comprising 5 or 1% of the area under the normal curve. In Fig. 9.4-2a, for $\alpha = 0.05$, the distance from 0 to Z_c is 1.645 (obtained from Appendix Table K) by the following reasoning. The probability α of making a Type I error is 0.05; therefore, the probability of correctly accepting the null is 0.95. The area under the curve below 0 is 0.500, and between 0 and Z_c it is 0.450 (see Appendix Table L). This probability (0.450) is equivalent to a Z_c (critical value) value of 1.645. Thus, any Z value greater than 1.645 is in the rejection region where $\alpha \leqslant 0.05$.

But suppose that the alternative hypothesis H_a is true, i.e., $\mu_1 - \mu_2 = \mu_D > \mu_o$. Then the distribution looks like that given in Fig. 9.4-2b. The area to the left of Z_c is the Type II error β, and the area to the right is the power of the test, hence power $= 1 - \beta$. In the form of a formula,

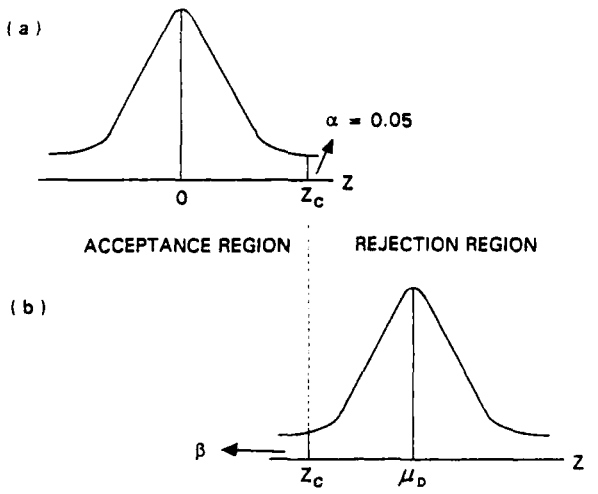


Fig. 9.4-2
Standard normal distribution of random variable Z with mean 0 and variance 1 illustrating the area corresponding to the α and β errors when the alternative hypothesis is actually true with mean μ_D.

$$\begin{aligned} \text{Power} &= P(Z > Z_c \text{ given } \mu_1 - \mu_2 = \mu_D) \\ &= 1 - \Phi[(Z_c - \mu_D)/SE] \end{aligned} \tag{9.4-3}$$

where Φ, denotes the cumulative distribution function of the standard normal random variable Z with mean zero and variance 1. The areas associated with the standard normal distribution are given in Appendix Table L. Examples for calculating the Type II error and consequently the power of the test will be given in the following sections.

Sensory Equivalence

Many claims substantiation in the personal care, household, and the food industries uses the human basic senses as the measuring instruments; specifically, the human basic senses of taste, sight, feel, and smell. In this section, a sensory equivalence is proposed to denote equivalence or parity in one or more of the basic senses between products. Obviously, sensory equivalence is used to support product parity. One unique application is the specification of a certain score on a rating scale that should be satisfied in a monadic evaluation for product acceptance.

During product development and reformulation, prototypes are monadically evaluated. It is common to establish a base point on the scale, such as a 6.0 on the 9-point hedonic scale to be the value of the null hypothesis:

$$H_o: \mu_o = 6.0$$

$$H_a: \mu_o \neq 6.0$$

Although the above formulation is two-sided, one may formulate the one-sided, H_a: $\mu_o < 6.0$ or H_a: $\mu_o > 6.0$, depending on the experimental problem.

Using the confidence interval test, if the interval of μ_o includes 6.0, then one may conclude sensory equivalence between the product and the perceived value of the null hypothesis. As discussed earlier, if more statistical evidence is required the power of the test can be computed or the number of tests (replication) increased and the result of each test plotted on a control chart (Fig. 9.4-1). The use of the control chart is useful because of the dynamic nature of consumer responses.

Example 9.4-1

Marketing research desires that the prototype of a product recently developed by R&D (Research and Development) should have an average score of at least 6.0 on the 9-point hedonic scale before further consumer testing can proceed. A monadic test with 100 panelists was conducted with the following result for "overall liking" of the product:

Mean	5.2
Standard deviation	1.4

In this example, H_o: $\mu_o \geqslant 6.0$ and H_a: $\mu_o < 6.0$. For large sample size ($N > 30$) one may use $Z_{\alpha/2}$ to approximate $t_{\alpha/2}$ in Eq. (9.4-1) for any α values. At the 95% confidence level, the value of $Z_{\alpha/2}$ is 1.960 (Table L). That is, the area $\alpha/2 = 0.5000 - 0.4750 = 0.025$ corresponds to $Z = 1.960$. Thus the 95% confidence interval is, from Eq. (9.4-1)

$$5.2 \pm 1.960(1.4/\sqrt{100}) = 5.2 \pm 0.27$$

resulting in an interval of 4.9 to 5.5, which can also be written as (4.9, 5.5). The interval does not include 6.0, therefore, one concludes that the prototype is not at parity with the perceived value of the null hypothesis. In fact, it is significantly inferior from the perceived value of the null at the $1.0 - 0.95 = 0.05$ significance level. Note that only in the confidence interval test can one perform this operation to obtain the level of significance. This operation should not be used in other test procedures involving more than two means, such as in multiple comparison tests, because the significance level increases with the number of means. A statistician should be consulted.

Let us calculate the power of the test when the observed result $\bar{X}$ is actually true; i.e., we reject the null and accept the alternative hypothesis. First we calculate the

critical value X_c on the $\bar{X}$ scale corresponding the cut-off point -1.960 on the Z scale. Since $Z = [(\bar{X} - 6.0)/(1.4/\sqrt{100}$, we find X_c to be $6.0 - 1.960(0.27) = 5.47$ (Fig. 9.4-3a). Then, using Table L to obtain the probability of Type II error, which is 0.159, we have using Eq. (9.4-3)

$$\begin{aligned}\text{Power} &= 1 - \Phi[(5.47 - 5.20)/0.27] \\ &= 1 - \Phi(1.00) = 1 - (0.5000 - 0.3413) \\ &= 1 - 0.159 \\ &= 0.841\end{aligned}$$

This result is reasonable because the difference between the null value of 6.0 and the alternative value of 5.2 is quite large; hence the power should be relatively large, suggesting the falsity of the null. See Fig. 9.4-3b for the graphical illustration of the result.

Suppose that the observed mean is 6.10 (Fig. 9.4-4b) instead of 5.20 (Fig. 9.4-3b). What is the power of the test? Note that the value of H_a is shifted to the right of H_o. Corresponding to the cutoff point 1.960 on the Z scale, the cutoff point on the $\bar{X}$ scale is

$$X_c = 6.0 + 1.96(0.27) = 6.53$$

and

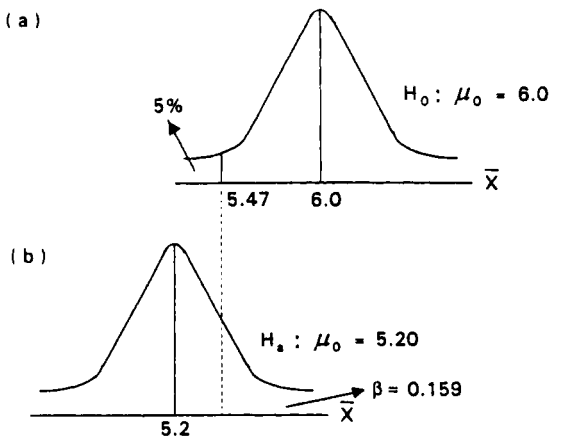


Fig. 9.4-3
Normal distribution curves for data from Example 9.4-1 showing the area of Type II error when the alternative hypothesis is true.

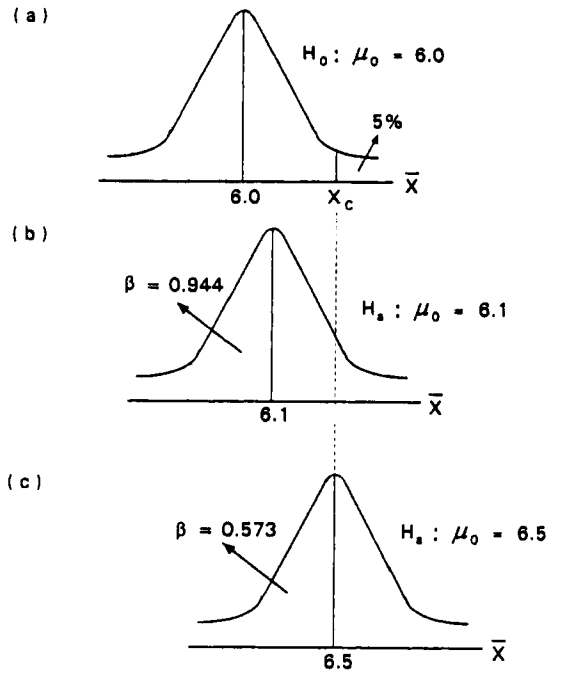


Fig. 9.4-4
Normal distribution curves for data from Example 9.4-1 showing the area of Type II error when the value of the alternative hypothesis is close to the null hypothesis.

$$\begin{aligned}\text{Power} &= 1 - \{1 - \Phi[(6.53 - 6.10)/0.27]\} \\ &= 1 - [1 - \Phi(1.593)] \\ &= 1 - [1 - (0.5000 - 0.4441)] \\ &= 1 - 0.944 = 0.056\end{aligned}$$

Again this result is reasonable because the value of 6.1 is close to null value of 6.0 and the power should be low (Fig. 9.4-4b) suggesting that the null is true. When the alternative is 6.50, the power is 0.457 (Fig. 9.4-4c). If we further increase the value of the alternative to 6.7, the power of the test becomes 0.764. This example clearly illustrates that as the value of H_a moves away farther from H_o the power of the test approaches 1.0.

Example 9.4-2

In a consumer study product A was compared to the leading brand product B. Using the paired comparison design the mean values for overall liking are as follows:

	A	B
No. of panelists, N	92	92
Mean, $\bar{X}$	6.1	5.9

The standard deviation of the difference is computed to be 1.10. Thus the standard error is SE = $1.10/\sqrt{92} = 0.11$. The 95% confidence interval of the difference is

$$(6.1 - 5.9) \pm 1.960(0.11) = 0.20 \pm 0.22$$

or an interval of $(-0.02, 0.42)$. Since the interval includes zero, one concludes at the 95% confidence level that products A and B are sensorially equivalent. Note that the lower limit of the confidence interval is close to zero, and one may suspect the parity or equivalency conclusion. Therefore the power of the test should be computed. In this example it is apparent that the value of the null is zero, hence the critical value X_c is 0.0 + 1.960(0.11) = 0.22. Note that X_c is to the right of μ_o (Fig. 9.4-5), therefore

$$\begin{aligned}
\text{Power} &= 1 - [1 - \Phi((0.22 - 0.20)/0.11)] \\
&= 1 - [1 - \Phi(0.182)] \\
&= 1 - [1 - (0.5000 - 0.0714)] \\
&= 1 - (1 - 0.428) \\
&= 0.428
\end{aligned}$$

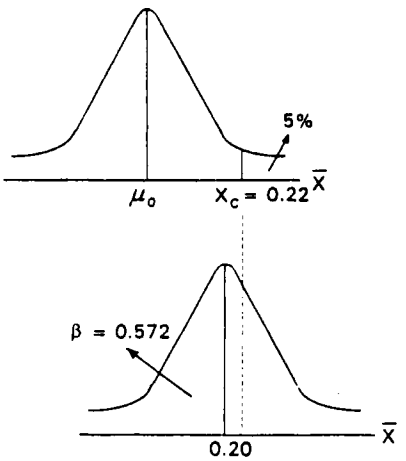


Fig. 9.4-5
Illustration of paired comparison test (Example 9.4-2) showing the area of Type II error β when power of test is low.

The power is less than 0.50, thus this result is a suspect to support parity. The conclusion of parity should be suspended until additional consumer test data are available.

Example 9.4-3

The results of four central location tests (CLT) and one conducted by R&D are given in Table 9.4-1. The question is: Is product A sensorially at parity with product B? The pooled standard error of difference for this study was found to be 0.123. The 99% confidence interval of the sensory equivalence zone is $3(0.123) = \pm 0.37$. This interval defines the entire width of the zone on the $\bar{d}$ chart.

Table 9.4-1
Overall liking mean scores for products A and B to illustrate the position of sensory equivalence.

Product	CLT1	CLT2	CLT3	CLT4	R&D
A	5.7±1.9	5.5±2.0	5.9±2.0	5.6±2.1	5.4±1.7
B	5.8±1.8	5.9±1.6	5.7±2.1	5.4±1.9	5.4±1.8
A–B	−0.1	−0.4	0.2	0.2	0.0
N	96	101	90	82	100

Note: R&D = Research and development personnel.

Using this interval, the plot of the difference A − B is shown in Fig. 9.4-6. It is seen that 4/5 of the differences lie in the zone of sensory equivalence, and for practical purposes this is an evidence of product parity. Perhaps, the data for CLT2 may be examined to see if this result is due to location difference (assignable cause) in product preference or strictly sampling variations (unassignable causes). The chart also could be interpreted to indicate that 1% of the time an average difference can be erroneously classified as out of the parity zone.

Sample Size and Power of the Test

As stated earlier, the number of observations N plays a major role in the power of the test by lowering the standard error of the mean. Recall that $SE = S/\sqrt{N}$ where S is the standard deviation of the observations. The formula for computing power of the test can also be written as

$$\text{Power} = 1 - \Phi[((\mu_1 - \mu_2)/SE) + Z_\alpha] \tag{9.4-4}$$

where Z_α is the standard normal deviation corresponding to the area of the normal curve at α. In the form of Eq. (9.4-4), one can vary N and fix α at certain level in computing the power.

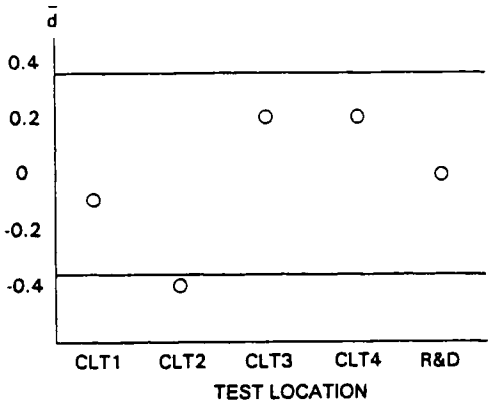


Fig. 9.4-6
Control chart to substantiate sensory equivalence between two products at the 95% confidence level. Note: The width of the zone is $\pm\sigma$ where σ estimated by the standard error of difference SE.

Let us use the statistics of Example 9.4-2 to illustrate the effect of N on power of the test. Using Eq. (9.4-4) the power was calculated for N = 41, 92, 184 with α equal to 0.05 and its corresponding Z_α = 1.960 (Table L). The result is shown in Table 9.4-2. The power of the test for rejecting a false null hypothesis increases as the value of the alternative hypothesis moves away from the null regardless of

Table 9.4-2
The effect of sample size on the power of the test with probability of Type I error fixed at 5%.

	Power		
$\mu_A - \mu_B$	N = 41	N = 92	N = 184
0.00	0.025	0.025	0.025
0.05	0.048	0.064	0.090
0.10	0.084	0.138	0.234
0.15	0.138	0.256	0.457
0.20	0.213	0.413	0.695
0.25	0.306	0.586	0.870
0.30	0.414	0.743	0.959
0.35	0.530	0.861	0.991
0.40	0.643	0.936	0.999
0.45	0.744	0.975	1.000
0.50	0.828	0.992	1.000

Note: This test is one-sided, $\alpha/2 = 0.05/2 = 0.025$, $Z_{\alpha/2} = 1.960$, standard deviation of difference = 1.10. $\mu_D = \mu_A - \mu_B$

the sample size. However, the power converges to 1.0 faster with larger sample sizes. For example, if $\mu_D = 0.45$ the power for N = 41, 92, 184 are, respectively, 0.744, 0.975, and 1.0. Note in this table that when the null hypothesis is true, $\mu_D = 0$, both the Type I error α and the power of the test have the same value equal to 0.025 regardless of N.

9.5 NULL HYPOTHESIS WITH SPECIFIED DIFFERENCE

One method of improving the support of parity is by specifying a difference in the formulation of the null hypothesis (Blackwelder 1982; Blackwelder and Chang 1984). For example, FDA (Federal Drug Administration) requires a 1-log reduction by the CADE handwashing procedure (Cade 1951) for antimicrobial efficacy. The null hypothesis may be written as

$$H_o: \mu_o - \mu_1 \geqslant 1 \text{ log}$$

and the alternative

$$H_a: \mu_o - \mu_1 < 1 \text{ log}$$

Here μ_o is the average microbial population at the beginning of the test (no treatment applied) and μ_1 is the average count after the treatment has been applied. The decision rule using the confidence interval test for the above hypotheses is as follows:

If the lower limit of the interval is equal to or greater than 1-log, one concludes that the required reduction for antimicrobial efficacy has been satisfied. On the other hand, if the interval includes a value less than 1 then the difference did not meet the 1-log reduction.

The choice of the specified difference depends on the experimental problem. The specification of the difference minimizes doubts in a claim that an inappropriate experimental design was used to support the acceptance of the null hypothesis. This specification also provides the appropriate sample size to obtain the desired power of the test.

Example 9.5-1

This example illustrates the computational procedure for testing the null hypothesis with a specified difference to guarantee product equivalence or superiority. The approach here is to decide how close the difference between products would be before one can guarantee that the products are sensorially equivalent. The sensory analyst will proceed to do a full central location test only if the small scale consumer test will result in a sensory equivalence with a specified difference $\mu_D = 0.25$ on the 9-point hedonic scale. In practice, this difference is close to 1 standard error generally observed in sensory evaluation work. For this example, let us use the partial data in Table 9.5-1 with 10 panelists.

The null and alternative hypotheses for this study are as follows:

$$H_o: \mu_1 \geqslant \mu_2 + 0.25$$

$$H_a: \mu_1 < \mu_2 + 0.25$$

Here, μ_1 is estimated by $\bar{X}_1$, the mean value for product 1, and μ_2 estimated by $\bar{X}_2$, the mean value for product 2. The estimated average difference between products under the null is $\bar{d} = 6.60 - 5.70 - 0.25 = 0.650$ with a standard error of difference (SE) of 0.314. Using Eq. (9.4-1), the 95% confidence interval is

$$0.650 \pm 2.262(0.314) = 0.650 \pm 0.710$$

or an interval of -0.06 to 1.36. Since this interval includes zero, the two products are sensorially equivalent. The sensory analyst had strong evidence that if a central location test is to be conducted there is a very good chance that product 1 should perform equally or even better than product 2 due to the inclusion of the specified difference in the null hypothesis.

What is the power of the test under the null hypothesis? From Eq. (9.4-4),

$$\text{Power} = 1 - \Phi[(.650/0.314) - 2.262]$$
$$= 1 - \Phi(-2.058) = 0.98$$

indicating that the null hypothesis is most likely true because of the low power for the alternative hypothesis, thus demonstrating the parity claim overwhelmingly.

Table 9.5-1
Sensory data for Example 9.5-1 to illustrate the computation for testing of null hypothesis with a specified difference.

Panelist	Product 1	Product 2	d
1	7	5	1.75
2	5	5	−0.25
3	8	6	1.75
4	6	7	−1.25
5	7	6	0.75
6	6	5	0.75
7	7	6	0.75
8	6	6	−0.25
9	7	6	0.75
10	7	5	1.75
Mean	6.60	5.70	0.65
Std. dev.	0.84	0.67	0.99

Note: d = (Product 1) − (Product 2) − 0.25

STATISTICAL TABLES

TABLE

A	The F distribution
B	Significant ranges for the Duncan's multiple range test
C	Critical ranges for the two-way classification comparing all possible pairs of items: Rank sum multiple comparison test
D	The t distribution
E	Critical values for testing extreme observations: The r ratio test
F	Critical values of the Grubbs ratio test for testing largest/smallest observation
G	Critical values of the Grubbs ratio test for testing two largest/two smallest observations
H	Critical values of the u test for testing largest/smallest observation with known standard deviation on v degrees of freedom
I	The Chi-square distribution
J	The cumulative normal distribution function: cumulative proportions to unit normal deviates
K	The upper percentile points for $q_{\alpha,t}$ for the range of independent normal random variable
L	Areas of the normal curve
M	Tables of balanced incomplete block designs for block size k = 2
N	Some popular Taguchi orthogonal array designs

Table A
The F distribution

5% Level

DF for denominator	DF for numerator									
	1	2	3	4	5	6	8	12	24	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	243.9	249.0	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
∞	3.84	2.99	2.60	2.37	2.21	2.10	1.94	1.75	1.52	1.00

Table A (cont.)

1% Level

DF for denominator	DF for numerator									
	1	2	3	4	5	6	8	12	24	∞
1	4052	4999	5403	5625	5764	5859	5982	6106	6234	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.37	99.42	99.46	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.49	27.05	26.60	26.12
4	21.20	18.00	16.69	15.98	15.52	15.21	14.80	14.37	13.93	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.29	9.89	9.47	9.02
6	13.74	10.92	9.78	9.15	8.75	8.47	8.10	7.72	7.31	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.84	6.47	6.07	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.67	5.28	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.47	5.11	4.73	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.06	4.71	4.33	3.91
11	9.65	7.20	6.22	5.67	5.32	5.07	4.74	4.40	4.02	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.16	3.78	3.36
13	9.07	6.70	5.74	5.20	4.86	4.62	4.30	3.96	3.59	3.16
14	8.86	6.51	5.56	5.03	4.69	4.46	4.14	3.80	3.43	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.00	3.67	3.29	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.55	3.18	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.79	3.45	3.08	2.65
18	8.28	6.01	5.09	4.58	4.25	4.01	3.71	3.37	3.00	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.30	2.92	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.23	2.86	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.51	3.17	2.80	2.36
22	7.94	5.72	4.82	4.31	3.99	3.76	3.45	3.12	2.75	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.41	3.07	2.70	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.36	3.03	2.66	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.32	2.99	2.62	2.17
26	7.72	5.53	4.46	4.14	3.82	3.59	3.29	2.96	2.58	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.26	2.93	2.55	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.23	2.90	2.52	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.20	2.87	2.49	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.84	2.47	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.66	2.29	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.82	2.50	2.12	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.66	2.34	1.95	1.38
∞	6.64	4.60	3.78	3.32	3.02	2.80	2.51	2.18	1.79	1.00

Source: Merrington, M. and Thompson, C. 1943. Tables of percentage points of the inverted beta(F) distribution. Biometrika 33:73–99. Reproduced with permission of the Biometrika Trustees.

Table B
Significant ranges for Duncan's multiple range test

	5% Level								
DF error	p = 2	3	4	5	6	7	8	9	10
1	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0
2	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09
3	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50
4	3.93	4.01	4.02	4.02	4.02	4.02	4.02	4.02	4.02
5	3.64	3.74	3.79	3.83	3.83	3.83	3.83	3.83	3.83
6	3.46	3.58	3.64	3.68	3.68	3.68	3.68	3.68	3.68
7	3.35	3.47	3.54	3.58	3.60	3.61	3.61	3.61	3.61
8	3.26	3.39	3.47	3.52	3.55	3.56	3.56	3.56	3.56
9	3.20	3.34	3.41	3.47	3.50	3.52	3.52	3.52	3.52
10	3.15	3.30	3.37	3.43	3.46	3.47	3.47	3.47	3.47
11	3.11	3.27	3.35	3.39	3.43	3.44	3.45	3.46	3.46
12	3.08	3.23	3.33	3.36	3.40	3.42	3.44	3.44	3.46
13	3.06	3.21	3.30	3.35	3.38	3.41	3.42	3.44	3.45
14	3.03	3.18	3.27	3.33	3.37	3.39	3.41	3.42	3.44
15	3.01	3.16	3.25	3.31	3.36	3.38	3.40	3.42	3.43
16	3.00	3.15	3.23	3.30	3.34	3.37	3.39	3.41	3.43
17	2.98	3.13	3.22	3.28	3.33	3.36	3.38	3.40	3.42
18	2.97	3.12	3.21	3.27	3.32	3.35	3.37	3.39	3.41
19	2.96	3.11	3.19	3.26	3.31	3.35	3.37	3.39	3.41
20	2.95	3.10	3.18	3.25	3.30	3.34	3.36	3.38	3.40
22	2.93	3.08	3.17	3.24	3.29	3.32	3.35	3.37	3.39
24	2.92	3.07	3.15	3.22	3.28	3.31	3.34	3.37	3.38
26	2.91	3.06	3.14	3.21	3.27	3.30	3.34	3.36	3.38
28	2.90	3.04	3.13	3.20	3.26	3.30	3.33	3.35	3.37
30	2.89	3.04	3.12	3.20	3.25	3.29	3.32	3.35	3.37
40	2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35
60	2.83	2.98	3.08	3.11	3.20	3.24	3.28	3.31	3.33
100	2.80	2.95	3.05	3.12	3.18	3.22	3.26	3.29	3.32
∞	2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.26	3.29

Table B (cont.)

1% Level

DF error	p = 2	3	4	5	6	7	8	9	10
1	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0
2	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0
3	8.26	8.5	8.6	8.7	8.8	8.9	8.9	9.0	9.0
4	6.51	6.8	6.9	7.0	7.1	7.1	7.2	7.2	7.3
5	5.70	5.96	6.11	6.18	6.26	6.33	6.40	6.41	6.3
6	5.24	5.51	5.65	5.73	5.81	5.88	5.95	6.00	6.0
7	4.95	5.22	5.37	5.45	5.53	5.61	5.69	5.73	5.8
8	4.74	5.00	5.14	5.23	5.32	5.40	5.47	5.51	5.5
9	4.60	4.86	4.99	5.08	5.17	5.25	5.32	5.36	5.4
10	4.48	4.73	4.88	4.96	5.06	5.13	5.20	5.24	5.28
11	4.39	4.63	4.77	4.86	4.94	5.01	5.06	5.12	5.15
12	4.32	4.55	4.68	4.76	4.84	4.92	4.96	5.02	5.07
13	4.26	4.48	4.62	4.69	4.74	4.84	4.88	4.94	4.98
14	4.21	4.42	4.55	4.63	4.70	4.78	4.83	4.87	4.91
15	4.17	4.37	4.50	4.58	4.64	4.72	4.77	4.81	4.84
16	4.13	4.34	4.45	4.54	4.60	4.67	4.72	4.76	4.79
17	4.10	4.30	4.41	4.50	4.56	4.63	4.68	4.72	4.75
18	4.07	4.27	4.38	4.46	4.53	4.59	4.64	4.68	4.71
19	4.05	4.24	4.35	4.43	4.50	4.56	4.61	4.64	4.67
20	4.02	4.22	4.33	4.40	4.47	4.53	4.58	4.61	4.65
22	3.99	4.17	4.28	4.36	4.42	4.48	4.53	4.57	4.60
24	3.96	4.14	4.24	4.33	4.39	4.44	4.49	4.53	4.57
26	3.93	4.11	4.21	4.30	4.36	4.41	4.46	4.50	4.53
28	3.91	4.08	4.18	4.28	4.34	4.39	4.43	4.47	4.51
30	3.89	4.06	4.16	4.22	4.32	4.36	4.41	4.45	4.48
40	3.82	3.99	4.10	4.17	4.24	4.30	4.34	4.37	4.41
60	3.76	3.92	4.03	4.12	4.17	4.23	4.27	4.31	4.34
100	3.71	3.86	3.98	4.06	4.11	4.17	4.21	4.25	4.29
∞	3.64	3.80	3.90	3.98	4.04	4.09	4.14	4.17	4.20

Source: Duncan, D.B. 1955. Multiple range and multiple F tests. Biometrics *11*:1-42. Reproduced with permission of The Biometric Society.

Table C
Critical ranges for the two-way classification comparing all possible pairs of items (treatments): Rank sum multiple comparison test

		Number of treatments												
N	α	3	4	5	6	7	8	9	10	11	12	13	14	15
5	.01	9	12	16	19	23	26	29	33	37	40	44	47	51
	.05	8	11	14	17	20	23	26	30	34	37	40	43	47
	.10	7	10	13	16	19	22	25	28	32	35	38	42	45
6	.01	10	14	17	21	25	29	33	37	41	45	49	53	57
	.05	9	12	15	19	22	26	29	33	37	41	43	48	52
	.10	8	11	14	17	20	24	27	31	34	38	42	45	49
7	.01	11	15	19	23	27	31	36	40	44	49	53	58	62
	.05	9	13	16	20	24	28	32	36	40	44	48	52	56
	.10	8	11	15	18	22	26	29	33	37	41	45	49	53
8	.01	12	16	20	25	29	34	38	43	47	52	57	62	67
	.05	10	14	17	21	25	30	34	38	42	47	51	56	60
	.10	9	12	16	20	24	27	31	36	40	44	48	52	57
9	.01	12	17	22	26	31	36	41	46	51	56	61	66	71
	.05	10	14	18	23	27	31	36	40	45	50	54	59	64
	.10	9	13	17	21	25	29	33	38	42	47	51	55	60
10	.01	13	18	23	28	33	38	43	49	54	59	65	70	75
	.05	11	15	19	24	28	33	38	43	47	52	57	62	67
	.10	9	13	18	22	26	31	35	40	44	49	54	58	63
11	.01	14	19	24	29	35	40	46	51	57	62	68	74	78
	.05	11	15	20	25	30	35	40	45	50	55	60	65	71
	.10	10	14	18	23	27	32	37	42	46	51	56	61	66
12	.01	14	20	25	31	36	42	48	54	59	65	71	77	83
	.05	12	16	21	26	31	36	41	47	52	58	63	68	74
	.10	10	15	19	24	29	33	38	43	48	54	59	64	69
13	.01	15	21	26	32	38	44	50	56	62	68	74	80	87
	.05	12	17	22	27	32	38	43	49	54	60	65	71	77
	.10	11	15	20	25	30	35	40	45	50	56	61	67	72
14	.01	16	21	27	33	39	45	52	58	64	71	77	84	90
	.05	13	17	23	28	34	39	45	50	56	62	68	74	80
	.10	11	16	21	26	31	36	41	47	52	58	63	69	75
15	.01	16	22	28	34	41	47	54	60	67	73	80	87	94
	.05	13	18	24	29	35	40	46	52	58	64	70	76	83
	.10	11	16	21	27	32	37	43	48	54	60	66	72	77
16	.01	16	23	29	36	42	49	56	62	69	76	83	90	97
	.05	13	18	24	30	36	42	48	54	60	67	73	79	86
	.10	12	17	22	27	33	38	44	50	56	62	68	74	80
17	.01	17	23	30	37	43	50	57	64	71	79	86	93	100
	.05	14	19	25	31	37	43	50	56	62	69	75	82	88
	.10	12	17	23	28	34	40	46	52	58	64	70	76	82

Table C (cont.)

N	α	3	4	5	6	7	8	9	10	11	12	13	14	15
18	.01	17	24	31	38	45	52	59	66	74	81	88	96	103
	.05	14	20	26	32	38	45	51	57	64	71	77	84	91
	.10	12	18	23	29	35	41	47	53	59	66	72	78	85
19	.01	18	25	32	39	46	53	61	68	76	83	91	98	106
	.05	14	20	27	33	39	46	52	59	66	73	80	86	93
	.10	13	18	24	30	36	42	48	54	61	67	74	80	87
20	.01	18	25	33	40	47	55	62	70	78	85	93	101	109
	.05	15	21	27	34	40	47	54	61	68	75	82	89	96
	.10	13	19	25	31	37	43	49	56	62	69	76	83	89
21	.01	19	26	33	41	48	56	64	72	79	87	95	104	112
	.05	15	21	28	35	41	48	55	62	69	76	84	91	98
	.10	13	19	25	31	38	44	51	57	64	71	78	85	92
22	.01	19	27	34	42	49	57	65	73	81	89	98	106	114
	.05	16	22	29	35	42	49	56	64	71	78	86	93	101
	.10	14	20	26	32	39	45	52	59	66	72	79	87	94
23	.01	20	27	35	43	51	59	67	75	83	91	100	108	117
	.05	16	22	29	36	43	50	58	65	72	80	88	95	103
	.10	14	20	26	33	39	46	53	60	67	74	81	89	96
24	.01	20	28	36	44	52	60	68	76	85	93	102	111	119
	.05	16	23	30	37	44	51	59	66	74	82	89	97	105
	.10	14	20	27	34	40	47	54	61	68	76	83	90	98
25	.01	21	28	36	44	53	61	70	78	87	95	104	113	122
	.05	17	23	31	38	45	52	60	68	75	83	91	99	107
	.10	15	21	27	34	41	48	55	63	70	77	85	92	100
26	.01	21	29	37	45	54	62	71	80	88	97	106	115	124
	.05	17	24	31	38	46	54	61	69	77	85	93	101	109
	.10	15	21	28	35	42	49	56	64	71	79	86	94	102
27	.01	21	30	38	46	55	63	72	81	90	99	108	117	127
	.05	17	24	32	39	47	55	62	70	78	87	95	103	111
	.10	15	22	29	36	43	50	57	65	73	80	88	96	104
28	.01	22	30	39	47	56	65	74	83	92	101	110	120	129
	.05	18	25	32	40	48	56	64	72	80	88	97	105	113
	.10	15	22	29	36	44	51	59	66	74	82	90	98	106
29	.01	22	31	39	48	57	66	75	84	93	103	112	122	131
	.05	18	25	33	41	49	57	65	73	81	90	98	107	116
	.10	16	23	30	37	44	52	60	67	75	83	91	99	108
30	.01	23	31	40	49	58	67	76	86	95	104	114	124	133
	.05	18	26	33	41	49	58	66	74	83	91	100	109	117
	.10	16	23	30	38	45	53	61	68	76	85	93	101	109
31	.01	23	32	41	50	59	68	77	87	97	106	116	126	136
	.05	18	26	34	42	50	58	67	75	84	93	102	110	119
	.10	16	23	31	38	46	54	62	70	78	86	94	103	111

Table C (cont.)

N	α	3	4	5	6	7	8	9	10	11	12	13	14	15
32	.01	23	32	41	50	60	69	79	88	98	108	118	128	138
	.05	19	27	35	43	51	59	68	77	85	94	103	112	121
	.10	16	24	31	39	47	54	63	71	79	87	96	104	113
33	.01	24	33	42	51	61	70	80	90	100	110	120	130	140
	.05	19	27	35	43	52	60	69	78	87	96	105	114	123
	.10	17	24	32	39	47	55	64	72	80	89	97	106	115
34	.01	24	33	42	52	61	71	81	91	101	111	121	132	142
	.05	19	27	36	44	53	61	70	79	88	97	106	116	125
	.10	17	24	32	40	48	56	64	73	81	90	99	108	117
35	.01	24	34	43	53	62	72	82	92	103	113	123	134	144
	.05	20	28	36	45	53	62	71	80	89	99	108	117	127
	.10	17	25	33	41	49	57	65	74	83	91	100	109	118
36	.01	25	34	44	53	63	73	83	94	104	114	125	136	146
	.05	20	28	37	45	54	63	72	81	91	100	109	119	129
	.10	17	25	33	41	49	58	66	75	84	93	102	111	120
37	.01	25	35	44	54	64	74	85	95	105	116	127	137	148
	.05	20	29	37	46	55	64	73	82	92	101	111	121	130
	.10	18	25	33	42	50	59	67	76	85	94	103	112	122
38	.01	25	35	45	55	65	75	86	96	107	113	128	139	150
	.05	20	29	38	46	56	65	74	84	93	103	112	122	132
	.10	18	26	34	42	51	59	68	77	86	95	104	114	123
39	.01	26	35	45	56	66	76	87	98	108	119	130	141	152
	.05	21	29	38	47	56	66	75	85	94	104	114	124	134
	.10	18	26	34	43	51	60	69	78	87	96	106	115	125
40	.01	26	36	46	56	67	77	88	99	110	121	132	143	154
	.05	21	30	39	48	57	66	76	86	95	105	115	125	136
	.10	18	26	35	43	52	61	70	79	88	98	107	117	126
41	.01	26	36	47	57	68	78	89	100	111	122	133	145	156
	.05	21	30	39	48	58	67	77	87	97	107	117	127	137
	.10	19	27	35	44	53	62	71	80	89	99	108	118	128
42	.01	27	37	47	58	68	79	90	101	112	124	135	146	158
	.05	21	30	40	49	58	68	78	88	98	108	118	129	139
	.10	19	27	36	44	53	62	72	81	91	100	110	120	129
43	.01	27	37	48	58	69	80	91	102	114	125	137	148	160
	.05	22	31	40	49	59	69	79	89	99	109	120	130	141
	.10	19	27	36	45	54	63	72	82	92	101	111	121	131
44	.01	27	38	48	59	70	81	92	104	115	127	138	150	162
	.05	22	31	40	50	60	70	80	90	100	111	121	132	142
	.10	19	28	36	45	55	64	73	83	93	102	112	122	133
45	.01	28	38	49	60	71	82	93	105	116	128	140	152	163
	.05	22	31	41	51	60	70	81	91	101	112	122	133	144
	.10	19	28	37	46	55	65	74	84	94	104	114	124	134

Table C (cont.)

N	α	3	4	5	6	7	8	9	10	11	12	13	14	15
46	.01	28	39	49	60	72	83	94	106	118	129	141	153	165
	.05	22	32	41	51	61	71	81	92	102	113	124	135	145
	.10	20	28	37	46	56	65	75	85	95	105	115	125	136
47	.01	28	39	50	61	72	84	95	107	119	131	143	155	167
	.05	23	32	42	52	62	72	82	93	104	114	125	136	147
	.10	20	29	38	47	56	66	76	86	96	106	116	127	137
48	.01	29	39	50	62	73	85	96	108	120	132	144	157	169
	.05	23	32	42	52	62	73	83	94	105	115	126	137	149
	.10	20	29	38	47	57	67	77	87	97	107	117	128	138
49	.01	29	40	51	62	74	86	97	109	121	134	146	158	171
	.05	23	33	43	53	63	73	84	95	106	117	128	139	150
	.10	20	29	38	48	58	67	77	88	98	108	119	129	140
50	.01	29	40	51	63	75	86	98	110	123	135	147	160	172
	.05	23	33	43	53	64	74	85	96	107	118	129	140	152
	.10	21	30	39	48	58	68	78	88	99	109	120	131	141
51	.01	29	41	52	64	75	87	99	112	124	136	149	161	174
	.05	24	33	44	54	64	75	86	97	108	119	130	142	153
	.10	21	30	39	49	59	69	79	89	100	110	121	132	143
52	.01	30	41	52	64	76	88	100	113	125	138	150	163	176
	.05	24	34	44	54	65	76	87	98	109	120	132	143	155
	.10	21	30	40	49	59	69	80	90	101	111	122	133	144
53	.01	30	41	53	65	77	89	101	114	126	139	152	164	177
	.05	24	34	44	55	66	76	87	99	110	121	133	144	156
	.10	21	30	40	50	60	70	80	91	102	112	123	134	145
54	.01	30	42	53	65	77	90	102	115	127	140	153	166	179
	.05	24	34	45	55	66	77	88	100	111	122	134	146	158
	.10	21	31	40	50	60	71	81	92	103	114	125	136	147
55	.01	31	42	54	66	78	91	103	116	129	141	154	168	181
	.05	25	35	45	56	67	78	89	100	112	124	135	147	159
	.10	22	31	41	51	61	71	82	93	104	115	126	137	148
56	.01	31	43	54	67	79	91	104	117	130	143	156	169	182
	.05	25	35	46	56	67	79	90	101	113	125	137	148	161
	.10	22	31	41	51	62	72	83	94	105	116	127	138	150
57	.01	31	43	55	67	80	92	105	118	131	144	157	171	184
	.05	25	35	46	57	68	79	91	102	114	126	138	150	162
	.10	22	32	42	52	62	73	83	94	105	117	128	139	151
58	.01	31	43	55	68	80	93	106	119	132	145	159	172	186
	.05	25	36	46	57	69	80	91	103	115	127	139	151	163
	.10	22	32	42	52	63	73	84	95	106	118	129	141	152
59	.01	32	44	56	68	81	94	107	120	133	147	160	174	187
	.05	25	36	47	58	69	81	92	104	116	128	140	152	165
	.10	22	32	42	53	63	74	85	96	107	119	130	142	153

Table C (cont.)

N	α	3	4	5	6	7	8	9	10	11	12	13	14	15
60	.01	32	44	56	69	82	95	108	121	134	148	161	175	189
	.05	26	36	47	58	70	81	93	105	117	129	141	154	166
	.10	22	32	43	53	64	75	86	97	108	120	131	143	155
61	.01	32	44	57	70	82	95	109	122	135	149	163	176	190
	.05	26	37	48	59	70	82	94	106	118	130	143	155	168
	.10	23	33	43	53	64	75	86	98	109	121	132	144	156
62	.01	32	45	57	70	83	96	110	123	137	150	164	178	192
	.05	26	37	48	59	71	83	95	107	119	131	144	156	169
	.10	23	33	43	54	65	76	87	98	110	122	133	145	157
63	.01	33	45	58	71	84	97	110	124	138	151	165	179	193
	.05	26	37	48	60	72	83	95	108	120	132	145	157	170
	.10	23	33	44	54	65	76	88	99	111	123	134	146	159
64	.01	33	45	58	71	84	98	111	125	139	153	167	181	195
	.05	27	38	49	60	72	84	96	108	121	133	146	159	172
	.10	23	33	44	55	66	77	88	100	112	124	136	148	160
65	.01	33	46	59	72	85	98	112	126	140	154	168	182	196
	.05	27	38	49	61	73	85	97	109	122	134	147	160	173
	.10	23	34	44	55	66	78	89	101	113	125	137	149	161
66	.01	33	46	59	72	86	99	113	127	141	155	169	184	198
	.05	27	38	50	61	73	85	98	110	123	135	148	161	174
	.10	24	34	45	56	67	78	90	102	113	126	138	150	162
67	.01	34	47	60	73	86	100	114	128	142	156	170	185	199
	.05	27	38	50	62	74	86	98	111	124	136	149	162	176
	.10	24	34	45	56	67	79	90	102	114	126	139	151	164
68	.01	34	47	60	73	87	101	115	129	143	157	172	186	201
	.05	27	39	50	62	74	87	99	112	124	137	150	164	177
	.10	24	34	45	56	68	79	91	103	115	127	140	152	165
69	.01	34	47	60	74	88	101	116	130	144	158	173	188	202
	.05	28	39	51	63	75	87	100	113	125	138	152	165	178
	.10	24	35	46	57	68	80	92	104	116	128	141	153	166
70	.01	34	48	61	74	88	102	116	131	145	160	174	189	204
	.05	28	39	51	63	75	88	101	113	126	139	153	166	179
	.10	24	35	46	57	69	81	92	105	117	129	142	154	167
71	.01	35	48	61	75	89	103	117	132	146	161	175	190	205
	.05	28	40	51	64	76	88	101	114	127	140	154	167	181
	.10	24	35	46	58	69	81	93	105	118	130	143	156	168
72	.01	35	48	62	76	89	104	118	132	147	162	177	192	207
	.05	28	40	52	64	76	89	102	115	128	141	155	168	182
	.10	25	35	47	58	70	82	94	106	119	131	144	157	170
73	.01	35	49	62	76	90	104	119	133	148	163	173	193	208
	.05	28	40	52	64	77	90	103	116	129	142	156	170	183
	.10	25	36	47	59	70	82	94	107	119	132	145	158	171

Table C (cont.)

N	α	3	4	5	6	7	8	9	10	11	12	13	14	15
74	.01	35	49	63	77	91	105	120	134	149	164	179	194	210
	.05	29	40	52	65	77	90	103	117	130	143	157	171	185
	.10	25	36	47	59	71	83	95	108	120	133	146	159	172
75	.01	36	49	63	77	91	106	120	135	150	165	180	196	211
	.05	29	41	53	65	78	91	104	117	131	144	158	172	186
	.10	25	36	48	59	71	83	96	108	121	134	147	160	173
76	.01	36	50	63	78	92	106	121	136	151	166	182	197	212
	.05	29	41	53	66	79	92	105	118	132	145	159	173	187
	.10	25	36	48	60	72	84	96	109	122	135	148	161	174
77	.01	36	50	64	78	93	107	122	137	152	167	183	198	214
	.05	29	41	54	66	79	92	105	119	132	146	160	174	188
	.10	25	37	48	60	72	84	97	110	123	136	149	162	175
78	.01	36	50	64	79	93	108	123	138	153	168	184	200	215
	.05	29	41	54	67	80	93	106	120	133	147	161	175	189
	.10	26	37	49	60	73	85	98	110	123	136	150	163	176
79	.01	37	51	65	79	94	109	124	139	154	170	185	201	217
	.05	29	42	54	67	80	93	107	120	134	148	162	176	191
	.10	26	37	49	61	73	86	98	111	124	137	151	164	178
80	.01	37	51	65	80	94	109	124	140	155	171	186	202	218
	.05	30	42	55	67	81	94	107	121	135	149	163	177	192
	.10	26	37	49	61	74	86	99	112	125	138	152	165	179
81	.01	37	51	66	80	95	110	125	141	156	172	187	203	219
	.05	30	42	55	68	81	94	108	122	136	150	164	179	193
	.10	26	38	49	62	74	87	100	113	126	139	153	166	180
82	.01	37	51	66	81	96	110	126	141	157	173	189	205	221
	.05	30	42	55	68	82	95	109	123	137	151	165	180	194
	.10	26	38	50	62	74	87	100	113	126	140	153	167	181
83	.01	38	52	66	81	96	111	127	142	158	174	190	206	222
	.05	30	43	56	69	82	96	109	123	138	152	166	181	195
	.10	26	38	50	62	75	88	101	114	127	141	154	168	182
84	.01	38	52	67	82	97	112	127	143	159	175	191	207	223
	.05	30	43	56	69	83	96	110	124	138	153	167	182	197
	.10	27	38	50	63	75	88	101	115	128	142	155	169	183
85	.01	38	52	67	82	97	113	128	144	160	176	192	208	225
	.05	31	43	56	70	83	97	111	125	139	154	168	183	198
	.10	27	39	51	63	76	89	102	115	129	142	156	170	184
86	.01	38	53	67	83	98	113	129	145	161	177	193	209	226
	.05	31	43	57	70	84	97	111	126	140	155	169	184	199
	.10	27	39	51	64	76	89	103	116	130	143	157	171	185
87	.01	38	53	68	83	98	114	130	146	162	178	194	211	227
	.05	31	44	57	70	84	98	112	126	141	155	170	185	200
	.10	27	39	51	64	77	90	103	117	130	144	158	172	186

Table C (cont.)

N	α	3	4	5	6	7	8	9	10	11	12	13	14	15
88	.01	39	53	68	83	99	115	130	146	163	179	195	212	229
	.05	31	44	57	71	85	98	113	127	142	156	171	186	201
	.10	27	39	52	64	77	90	104	117	131	145	159	173	187
89	.01	39	54	69	84	99	115	131	147	164	180	196	213	230
	.05	31	44	58	71	85	99	113	128	142	157	172	187	202
	.10	27	39	52	65	78	91	104	118	132	146	160	174	189
90	.01	39	54	69	84	100	116	132	148	164	181	198	214	231
	.05	31	44	58	72	85	100	114	129	143	158	173	188	203
	.10	28	40	52	65	78	91	105	119	132	147	161	175	190
91	.01	39	54	69	85	101	117	133	149	165	182	199	215	232
	.05	32	45	58	72	86	100	115	129	144	159	174	189	205
	.10	28	40	52	65	78	92	105	119	133	147	162	176	191
92	.01	40	55	70	85	101	117	133	150	166	183	200	217	234
	.05	32	45	59	72	86	101	115	130	145	160	175	190	206
	.10	28	40	53	66	79	92	106	120	134	148	163	177	192
93	.01	40	55	70	86	102	118	134	151	167	184	201	218	235
	.05	32	45	59	73	87	101	116	131	146	161	176	191	207
	.10	28	40	53	66	79	93	107	121	135	149	163	178	193
94	.01	40	55	71	86	102	118	135	151	168	185	202	219	236
	.05	32	45	59	73	87	102	116	131	146	162	177	192	208
	.10	28	41	53	66	80	93	107	121	135	150	164	179	194
95	.01	40	55	71	87	103	119	136	152	169	186	203	220	237
	.05	32	46	59	73	88	102	117	132	147	162	178	193	209
	.10	28	41	54	67	80	94	108	122	136	151	165	180	193
96	.01	40	56	71	87	103	120	136	153	170	187	204	221	239
	.05	32	46	60	74	88	103	118	133	148	163	179	194	210
	.10	28	41	54	67	81	94	108	122	137	151	166	181	196
97	.01	41	56	72	88	104	120	137	154	171	188	205	222	240
	.05	33	46	60	74	89	103	118	133	149	164	180	195	211
	.10	29	41	54	67	81	95	109	123	138	152	167	182	197
98	.01	41	56	72	88	104	121	138	155	172	189	206	224	241
	.05	33	46	60	75	89	104	119	134	149	165	181	196	212
	.10	29	41	54	68	81	95	109	124	138	153	168	183	198
99	.01	41	57	72	89	105	122	138	155	172	190	207	225	242
	.05	33	47	61	75	90	104	120	135	150	166	182	197	213
	.10	29	42	55	68	82	96	110	124	139	154	169	184	199
100	.01	41	57	73	89	105	122	139	156	173	191	208	226	244
	.05	33	47	61	75	90	105	120	135	151	167	182	198	214
	.10	29	42	55	68	82	96	111	125	140	154	169	185	200
105	.01	42	58	75	91	108	125	143	160	178	195	213	231	250
	.05	34	48	63	77	92	108	123	139	155	171	187	203	220
	.10	30	43	56	70	84	99	113	128	143	158	174	189	205
110	.01	43	60	76	93	111	128	146	164	182	200	218	237	256
	.05	35	49	64	79	94	110	126	142	158	175	191	208	225
	.10	30	44	58	72	86	101	116	131	146	162	178	194	210

Source: Dunn-Rankin, Peter. 1965. The true probability distribution of the range of rank totals and its application to psychological scaling. Ph.D. Dissertation, The Florida State University, Tallahassee, Florida. Reproduced with permission of the author.

Table D
The t distribution

DF	.50	.40	.30	.20	.10	.05	.02	.01	.001
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.598
3	.765	.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	.741	.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	.727	.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	.718	.906	1.134	1.440	1.943	2.447	3.142	3.707	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.500	2.807	3.767
24	.685	.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	.681	.851	1.050	1.303	1.684	2.201	2.423	2.704	3.551
60	.679	.848	1.046	1.296	1.671	2.000	2.390	2.660	3.460
120	.677	.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373
∞	.674	.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

Source: Table generated using a SAS program written by R.W. Washam II, The Dial Technical and Administrative Center, Scottsdale, Arizona.

Table E
Critical values for testing extreme observations: The r ratio test

Statistic	n	Critical values						
		.30	20	.10	.05	.02	.01	.005
	3	.684	.781	.886	.941	.976	.988	.994
	4	.471	.560	.679	.765	.846	.889	.926
r_{10}	5	.373	.451	.557	.642	.729	.780	.821
	6	.318	.386	.482	.560	.644	.698	.740
	7	.281	.344	.434	.507	.586	.637	.680
	8	.318	.385	.479	.554	.631	.683	.725
r_{11}	9	.288	.352	.441	.512	.587	.635	.677
	10	.265	.325	.409	.477	.551	.597	.639
	11	.391	.442	.517	.576	.638	.679	.713
r_{21}	12	.370	.419	.490	.546	.605	.642	.675
	13	.351	.399	.467	.521	.578	.615	.649
	14	.370	.421	.492	.546	.602	.641	.674
	15	.353	.402	.472	.525	.579	.616	.647
	16	.338	.386	.454	.507	.559	.595	.624
r_{22}	17	.325	.373	.438	.490	.542	.577	.605
	18	.314	.361	.424	.475	.527	.561	.589
	19	.304	.350	.412	.462	.514	.547	.575
	20	.295	.340	.401	.450	.502	.535	.562
	21	.287	.331	.391	.440	.491	.524	.551
	22	.280	.323	.382	.430	.481	.514	.541
	23	.274	.316	.374	.421	.472	.505	.532
	24	.268	.310	.367	.413	.464	.497	.524
	25	.262	.304	.360	.406	.457	.489	.516

$$r_{10} = (X_2 - X_1)/(X_k - X_1)$$
$$r_{11} = (X_2 - X_1)/(X_{k-1} - X_1)$$
$$r_{21} = (X_3 - X_1)/(X_{k-1} - X_1)$$
$$r_{22} = (X_3 - X_1)/(X_{k-2} - X_1)$$

Source: Dixon, W.J. 1953. Processing data for outliers. Biometrics 9:74–89. Reproduced with permission of the Biometric Society.

Table F
Critical values of the Grubbs ratio test for testing largest/smallest observation

n	1%	2.5%	5%	10%
3	.0001	.0007	.0027	.0109
4	.0100	.0248	.0494	.0975
5	.0442	.0808	.1270	.1984
6	.0928	.1453	.2032	.2826
7	.1447	.2066	.2696	.3503
8	.1948	.2616	.3261	.4050
9	.2411	.3101	.3742	.4502
10	.2831	.3526	.4154	.4881
11	.3211	.3901	.4511	.5204
12	.3554	.4232	.4822	.5483
13	.3864	.4528	.5097	.5727
14	.4145	.4792	.5340	.5942
15	.4401	.5030	.5559	.6134
16	.4634	.5246	.5755	.6306
17	.4848	.5442	.5933	.6461
18	.5044	.5621	.6095	.6601
19	.5225	.5785	.6243	.6730
20	.5393	.5937	.6379	.6848
21	.5548	.6076	.6504	.6958
22	.5692	.6206	.6621	.7058
23	.5827	.6327	.6728	.7151
24	.5953	.6439	.6829	.7238
25	.6071	.6544	.6923	.7319

Source: Grubbs, F.E. 1950. Sample criteria for testing outlying observations. Annals Math. Stat. *21*:27–58. Reproduced with permission.

Table G
Critical values of the Grubbs ratio test for testing two largest/two smallest observation

n	1%	2.5%	5%	10%
4	.0000	.0002	.0008	.0031
5	.0035	.0090	.0183	.0376
6	.0186	.0349	.0565	.0921
7	.0440	.0708	.1020	.1479
8	.0750	.1101	.1478	.1994
9	.1082	.1492	.1909	.2454
10	.1415	.1865	.2305	.2863
11	.1736	.2212	.2666	.3226
12	.2044	.2536	.2996	.3552
13	.2333	.2836	.3295	.3843
14	.2605	.3112	.3568	.4106
15	.2859	.3367	.3818	.4345
16	.3098	.3603	.4048	.4562
17	.3321	.3822	.4259	.4761
18	.3530	.4025	.4455	.4944
19	.3725	.4214	.4636	.5113
20	.3909	.4391	.4804	.5269

Source: Grubbs, F.E. 1950. Sample criteria for testing outlying observations. Annals Math. Stat. *21*: 27-58. Reproduced with permission.

Table H
Critical values of the u test for testing largest/smallest observation with known standard deviation on v degrees of freedom

v	10% points						
DF	n = 3	4	5	6	7	8	9
10	1.68	1.93	2.11	2.25	2.36	2.46	2.54
11	1.66	1.91	2.08	2.21	2.32	2.42	2.49
12	1.65	1.89	2.05	2.19	2.29	2.38	2.46
13	1.63	1.87	2.04	2.16	2.27	2.36	2.43
14	1.62	1.85	2.02	2.14	2.25	2.33	2.41
15	1.61	1.84	2.00	2.13	2.23	2.31	2.39
16	1.61	1.83	1.99	2.12	2.22	2.30	2.37
17	1.60	1.82	1.98	2.10	2.20	2.28	2.35
18	1.59	1.82	1.97	2.09	2.19	2.27	2.34
19	1.59	1.81	1.97	2.09	2.18	2.26	2.33
20	1.58	1.80	1.96	2.08	2.17	2.25	2.32
24	1.57	1.78	1.94	2.05	2.15	2.22	2.29
30	1.55	1.77	1.92	2.03	2.12	2.20	2.26
40	1.54	1.75	1.89	2.01	2.09	2.17	2.23
60	1.52	1.73	1.87	1.98	2.07	2.14	2.20
120	1.51	1.71	1.85	1.96	2.05	2.12	2.18
∞	1.50	1.70	1.83	1.94	2.02	2.09	2.15

Table H (cont.)

v	5% points						
DF	n = 3	4	5	6	7	8	9
10	2.02	2.29	2.49	2.63	2.75	2.85	2.93
11	1.99	2.26	2.44	2.58	2.70	2.79	2.87
12	1.97	2.22	2.40	2.54	2.65	2.75	2.83
13	1.95	2.20	2.38	2.51	2.62	2.71	2.79
14	1.93	2.18	2.35	2.48	2.59	2.68	2.76
15	1.92	2.16	2.33	2.46	2.56	2.65	2.73
16	1.90	2.14	2.31	2.44	2.54	2.63	2.70
17	1.89	2.13	2.30	2.42	2.52	2.61	2.68
18	1.88	2.12	2.28	2.41	2.51	2.59	2.66
19	1.87	2.11	2.27	2.39	2.49	2.58	2.65
20	1.87	2.10	2.26	2.38	2.48	2.56	2.63
24	1.84	2.07	2.23	2.35	2.44	2.52	2.59
30	1.82	2.04	2.20	2.31	2.40	2.48	2.55
40	1.80	2.02	2.17	2.28	2.37	2.44	2.51
60	1.78	1.99	2.14	2.25	2.33	2.41	2.47
120	1.76	1.97	2.11	2.21	2.30	2.37	2.43
∞	1.74	1.94	2.08	2.18	2.27	2.33	2.39

Table H (cont.)

v	2.5% points						
DF	n = 3	4	5	6	7	8	9
10	2.36	2.64	2.84	2.99	3.10	3.20	3.28
11	2.31	2.59	2.78	2.93	3.04	3.14	3.22
12	2.28	2.55	2.74	2.88	2.99	3.08	3.16
13	2.25	2.52	2.70	2.84	2.95	3.04	3.12
14	2.23	2.49	2.67	2.80	2.91	3.00	3.08
15	2.20	2.46	2.64	2.77	2.88	2.97	3.04
16	2.19	2.44	2.62	2.75	2.85	2.94	3.01
17	2.17	2.42	2.60	2.73	2.83	2.92	2.99
18	2.16	2.41	2.58	2.71	2.81	2.89	2.97
19	2.15	2.39	2.56	2.69	2.79	2.87	2.95
20	2.14	2.38	2.55	2.67	2.77	2.86	2.93
24	2.10	2.34	2.50	2.62	2.72	2.80	2.87
30	2.07	2.30	2.46	2.58	2.67	2.75	2.82
40	2.04	2.27	2.42	2.53	2.62	2.70	2.76
60	2.01	2.23	2.38	2.49	2.58	2.65	2.71
120	1.98	2.20	2.34	2.45	2.53	2.60	2.66
∞	1.95	2.16	2.30	2.41	2.49	2.56	2.61

Table H (cont.)

v	1% points						
DF	n = 3	4	5	6	7	8	9
10	2.76	3.05	3.25	3.39	3.50	3.59	3.67
11	2.71	3.00	3.19	3.33	3.44	3.53	3.61
12	2.67	2.95	3.14	3.28	3.39	3.48	3.55
13	2.63	2.91	3.10	3.24	3.34	3.43	3.51
14	2.60	2.87	3.06	3.20	3.30	3.39	3.47
15	2.57	2.84	3.02	3.16	3.27	3.35	3.43
16	2.55	2.81	3.00	3.13	3.24	3.32	3.39
17	2.52	2.79	2.97	3.10	3.21	3.29	3.36
18	2.50	2.77	2.95	3.08	3.18	3.27	3.34
19	2.49	2.75	2.92	3.06	3.16	3.24	3.31
20	2.47	2.73	2.91	3.04	3.14	3.22	3.29
24	2.43	2.68	2.85	2.97	3.07	3.15	3.22
30	2.38	2.62	2.79	2.91	3.01	3.08	3.15
40	2.34	2.57	2.73	2.85	2.94	3.02	3.08
60	2.30	2.52	2.68	2.79	2.88	2.95	3.01
120	2.25	2.48	2.62	2.73	2.82	2.89	2.95
∞	2.22	2.43	2.57	2.68	2.76	2.83	2.88

Table H (cont.)

v	0.5% points						
DF	n = 3	4	5	6	7	8	9
10	3.03	3.32	3.52	3.65	3.76	3.85	3.92
11	2.98	3.27	3.46	3.60	3.71	3.79	3.86
12	2.93	3.22	3.41	3.55	3.66	3.74	3.81
13	2.89	3.18	3.37	3.50	3.61	3.70	3.77
14	2.86	3.14	3.33	3.46	3.57	3.66	3.73
15	2.83	3.11	3.29	3.43	3.53	3.62	3.69
16	2.80	3.08	3.26	3.39	3.50	3.58	3.66
17	2.77	3.05	3.23	3.36	3.47	3.55	3.62
18	2.75	3.02	3.21	3.34	3.44	3.53	3.60
19	2.73	3.00	3.18	3.31	3.42	3.50	3.57
20	2.72	2.98	3.16	3.29	3.39	3.48	3.55
24	2.66	2.92	3.09	3.22	3.32	3.40	3.47
30	2.60	2.86	3.02	3.15	3.25	3.32	3.39
40	2.55	2.79	2.96	3.08	3.17	3.25	3.31
60	2.50	2.73	2.89	3.01	3.10	3.17	3.23
120	2.45	2.67	2.83	2.94	3.02	3.09	3.15
∞	2.40	2.62	2.76	2.87	2.95	3.02	3.07

Table H (cont.)

v	0.1% points						
DF	n = 3	4	5	6	7	8	9
10	3.54	3.84	4.04	4.17	4.28	4.35	4.40
11	3.49	3.80	3.99	4.12	4.23	4.30	4.36
12	3.45	3.75	3.94	4.07	4.19	4.26	4.31
13	3.41	3.71	3.90	4.03	4.14	4.22	4.28
14	3.38	3.67	3.86	4.00	4.10	4.18	4.24
15	3.35	3.64	3.83	3.96	4.06	4.15	4.21
16	3.32	3.51	3.80	3.93	4.03	4.12	4.18
17	3.29	3.58	3.77	3.90	4.00	4.09	4.15
18	3.27	3.55	3.74	3.88	3.98	4.06	4.12
19	3.25	3.53	3.72	3.85	3.95	4.03	4.10
20	3.23	3.51	3.70	3.83	3.93	4.01	4.08
24	3.16	3.44	3.62	3.75	3.85	3.93	4.00
30	3.08	3.36	3.53	3.66	3.76	3.84	3.90
40	3.01	3.27	3.44	3.57	3.66	3.74	3.81
60	2.93	3.19	3.35	3.47	3.56	3.64	3.70
120	2.85	3.10	3.26	3.37	3.46	3.53	3.59
∞	2.78	3.01	3.17	3.28	3.36	3.43	3.48

Source: Nair, K.R. 1952. Tables of percentage points of the studentized extreme deviate from the sample mean. Biometrika *31*: 189–191. Reproduced with permission.

Table I
The Chi-square distribution

DF	.50	.30	.20	.10	.05	.02	.01	.001
1	0.455	1.074	1.642	2.706	3.841	5.412	6.635	10.827
2	1.386	2.408	3.219	4.605	5.991	7.824	9.210	13.815
3	2.366	3.665	4.642	6.251	7.815	9.837	11.345	16.266
4	3.357	4.878	5.989	7.779	9.488	11.668	13.277	18.467
5	4.351	6.064	7.289	9.236	11.070	13.388	15.086	20.515
6	5.348	7.231	8.558	10.645	12.592	15.033	16.812	22.457
7	6.346	8.383	9.803	12.017	14.067	16.622	18.475	24.322
8	7.344	9.524	11.030	13.362	15.507	18.168	20.090	26.125
9	8.343	10.656	12.242	14.684	16.919	19.679	21.666	27.877
10	9.342	11.781	13.442	15.987	18.307	21.161	23.209	29.588
11	10.341	12.899	14.631	17.275	19.675	22.618	24.725	31.264
12	11.340	14.011	15.812	18.549	21.026	24.054	26.217	32.909
13	12.340	15.119	16.985	19.812	22.362	25.472	27.688	34.528
14	13.339	16.222	18.151	21.064	23.685	26.873	29.141	36.123
15	14.339	17.322	19.311	22.307	24.996	28.259	30.578	37.697
16	15.338	18.418	20.465	23.542	26.296	29.633	32.000	39.252
17	16.338	19.511	21.615	24.769	27.587	30.995	33.409	40.790
18	17.338	20.601	22.760	25.989	28.869	32.346	34.805	42.312
19	18.338	21.689	23.900	27.204	30.144	33.687	36.191	43.820
20	19.337	22.775	25.038	28.412	31.410	35.020	37.566	45.315
21	20.337	23.858	26.171	29.615	32.671	36.343	38.932	46.797
22	21.337	24.939	27.301	30.813	33.924	37.659	40.289	48.268
23	22.337	26.018	28.429	32.007	35.172	38.968	41.638	49.728
24	23.337	27.096	29.553	33.196	36.415	40.270	42.980	51.179
25	24.337	28.172	30.675	34.382	37.652	41.566	44.314	52.620
26	25.336	29.246	31.795	35.563	38.885	42.856	45.642	54.052
27	26.336	30.319	32.912	36.741	40.113	44.140	46.963	55.476
28	27.336	31.391	34.027	37.916	41.337	45.419	43.278	56.893
29	28.336	32.461	35.139	39.087	42.557	46.693	49.588	58.302
30	29.336	33.530	36.250	40.256	43.773	47.962	50.892	59.703

Source: Table generated using a SAS program written by R.W. Washam II, The Dial Technical and Administrative Center, Scottsdale, Arizona.

Table J
The cumulative normal distribution function; cumulative proportions to unit normal deviates

	Decimal fraction of P(i,j)									
P(i,j), %	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
50	0.000	0.003	0.005	0.008	0.010	0.013	0.015	0.018	0.020	0.023
51	0.025	0.028	0.030	0.033	0.035	0.038	0.040	0.043	0.045	0.048
52	0.050	0.053	0.055	0.058	0.060	0.063	0.065	0.068	0.070	0.073
53	0.075	0.078	0.080	0.083	0.085	0.088	0.090	0.093	0.095	0.098
54	0.100	0.103	0.105	0.108	0.111	0.113	0.116	0.118	0.121	0.123
55	0.126	0.128	0.131	0.133	0.136	0.138	0.141	0.143	0.146	0.148
56	0.151	0.154	0.156	0.159	0.161	0.164	0.166	0.169	0.171	0.174
57	0.176	0.179	0.181	0.184	0.187	0.189	0.192	0.194	0.197	0.199
58	0.202	0.204	0.207	0.210	0.212	0.215	0.217	0.220	0.222	0.225
59	0.228	0.230	0.233	0.235	0.238	0.240	0.243	0.246	0.248	0.251
60	0.253	0.256	0.259	0.261	0.264	0.266	0.269	0.272	0.274	0.277
61	0.279	0.282	0.285	0.287	0.290	0.292	0.295	0.298	0.300	0.303
62	0.305	0.308	0.311	0.313	0.316	0.319	0.321	0.324	0.327	0.329
63	0.332	0.335	0.337	0.340	0.342	0.345	0.348	0.350	0.353	0.356
64	0.358	0.361	0.364	0.366	0.369	0.372	0.375	0.377	0.380	0.383
65	0.385	0.388	0.391	0.393	0.396	0.399	0.402	0.404	0.407	0.410
66	0.412	0.415	0.418	0.421	0.423	0.426	0.429	0.432	0.434	0.437
67	0.440	0.443	0.445	0.448	0.451	0.454	0.457	0.459	0.462	0.465
68	0.468	0.470	0.473	0.476	0.479	0.482	0.485	0.487	0.490	0.493
69	0.496	0.499	0.502	0.504	0.507	0.510	0.513	0.516	0.519	0.522
70	0.524	0.527	0.530	0.533	0.536	0.539	0.542	0.545	0.548	0.550
71	0.553	0.556	0.559	0.562	0.565	0.568	0.571	0.574	0.577	0.580
72	0.583	0.586	0.589	0.592	0.595	0.598	0.601	0.604	0.607	0.610
73	0.613	0.616	0.619	0.622	0.625	0.628	0.631	0.634	0.637	0.640
74	0.643	0.646	0.650	0.653	0.656	0.659	0.662	0.665	0.668	0.671
75	0.674	0.678	0.681	0.684	0.687	0.690	0.693	0.697	0.700	0.703

Table J (cont.)
The cumulative normal distribution function: cumulative proportions to unit normal deviates

P(i,j),%	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
76	0.706	0.710	0.713	0.716	0.719	0.722	0.726	0.729	0.732	0.736
77	0.739	0.742	0.745	0.749	0.752	0.755	0.759	0.762	0.765	0.769
78	0.772	0.776	0.779	0.782	0.786	0.789	0.793	0.796	0.800	0.803
79	0.806	0.810	0.813	0.817	0.820	0.824	0.827	0.831	0.834	0.838
80	0.842	0.845	0.849	0.852	0.856	0.860	0.863	0.867	0.871	0.874
81	0.878	0.882	0.885	0.889	0.893	0.896	0.900	0.904	0.908	0.912
82	0.915	0.919	0.923	0.927	0.931	0.935	0.938	0.942	0.946	0.950
83	0.954	0.958	0.962	0.966	0.970	0.974	0.978	0.982	0.986	0.990
84	0.994	0.999	1.003	1.007	1.011	1.015	1.019	1.024	1.028	1.032
85	1.036	1.041	1.045	1.049	1.054	1.058	1.063	1.067	1.071	1.076
86	1.080	1.085	1.089	1.094	1.098	1.103	1.108	1.112	1.117	1.122
87	1.126	1.131	1.136	1.141	1.146	1.150	1.155	1.160	1.165	1.170
88	1.175	1.180	1.185	1.190	1.195	1.200	1.206	1.211	1.216	1.221
89	1.227	1.232	1.237	1.243	1.248	1.254	1.259	1.265	1.270	1.276
90	1.282	1.287	1.293	1.299	1.305	1.311	1.317	1.323	1.329	1.335
91	1.341	1.347	1.353	1.359	1.366	1.372	1.379	1.385	1.392	1.398
92	1.405	1.412	1.419	1.426	1.433	1.440	1.447	1.454	1.461	1.468
93	1.476	1.483	1.491	1.499	1.506	1.514	1.522	1.530	1.538	1.546
94	1.555	1.563	1.572	1.580	1.589	1.598	1.607	1.616	1.626	1.635
95	1.645	1.655	1.665	1.675	1.685	1.695	1.706	1.717	1.728	1.739
96	1.751	1.762	1.774	1.787	1.799	1.812	1.825	1.838	1.852	1.866
97	1.881	1.896	1.911	1.927	1.943	1.960	1.977	1.995	2.014	2.034
98	2.054	2.075	2.097	2.120	2.144	2.170	2.197	2.226	2.257	2.290
99	2.326	2.366	2.409	2.457	2.512	2.576	2.652	2.748	2.878	3.090

Source: Table generated by a SAS program written by R.W. Washam II, The Dial Technical and Administrative Center, Scottsdale, Arizona.

Table K
The upper percentile points for $q_{\alpha,t}$ for the range of independent normal random variable

	Number of treatments												
α	3	4	5	6	7	8	9	10	11	12	13	14	15
0.01	0.19	0.43	0.66	0.87	1.05	1.20	1.34	1.47	1.58	1.68	1.77	1.86	1.93
0.05	0.43	0.76	1.03	1.25	1.44	1.60	1.74	1.86	1.97	2.07	2.16	2.24	2.32

Source: Pearson, E.S. and Hartley, H.O. 1954. Biometrika Tables for Statisticians, Vol. 1, Cambridge University Press. Reproduced with permission of the Biometrika Trustees.

Table L
Areas of the normal curve

	Decimal fraction of Z									
Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441

1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Source: Table generated using a SAS program written by R.W. Washam II, The Dial Technical and Administrative Center, Scottsdale, Arizona.

Table M
Tables of balanced incomplete block designs for block size k = 2

Some useful guides in designing a consumer test using the balanced incomplete block designs is given in this table. The terms in the formulas are defined in Chapter 4. In addition, q is always equal to 2 to account for order of sample presentation, i.e., AB, BA. The term p is the number of repetitions of the basic design. In practice, the size of p is determined by the desired number of judgments per sample, the desired number of judgments per pair, and cost consideration especially for large t. The order of product use as well as assignment of product pairs to the panelist were done at random using a SAS program. Products are denoted by letters A, B, C, etc.

Basic design: $t = 3$, $b = t(t-1) = 3$, $r = t - 1 = 2$,
$\lambda = 1$, $q = 2$, $p = 25$

$pbq = 25(3)2 = 150$ panelists are needed plus dropout/nonresponse allowance

$prq = 25(2)2 = 100$ judgments per sample

$p\lambda q = 25(1)2 = 50$ judgments per pair

Design: Repeat design in multiples of $bq = 6$

Panelist	Order of Product Use
1	AC
2	BC
3	AB
4	BA
5	CB
6	CA

Basic design: $t = 4$, $b = 6$, $r = 3$, $\lambda = 1$
$q = 2$, $p = 25$

$pbq = 25(6)2 = 300$ panelists

$prq = 25(3)2 = 150$ judgments per sample

$p\lambda 2 = 25(1)2 = 50$ judgments per pair

Design: Repeat design in multiples of $bq = 12$

Panelist	Order of Product Use
1	DC
2	BD
3	DA
4	AD
5	CB
6	AC
7	DB
8	AB
9	CD
10	BA
11	CA
12	BC

Table M (cont.)

Basic design: $t = 5, b = 10, r = 4, \lambda = 1$
$q = 2, p = 25$

$pbq = 25(10)2 = 500$ panelists

$prq = 25(4)2 = 200$ judgments per sample

$p\lambda q = 25(1)2 = 50$ judgment per pair

Design: Repeat design in multiples of $bq = 20$

Panelist	Order of Product Use
1	CA
2	AB
3	AC
4	BE
5	EB
6	BD
7	EA
8	CD
9	CE
10	DB
11	DA
12	CB
13	BC
14	AE
15	DC
16	AD
17	EC
18	DE
19	BA
20	ED

Basic design: $t = 6, b = 15, r = 5, \lambda = 1$
$q = 2, p = 20$

$pbq = 20(15)2 = 600$ panelists

$prq = 20(5)2 = 200$ judgments per sample

$pq = 20(1)1 = 40$ judgments per pair

Design: Repeat design in multiples of $bq = 30$

Panelist	Order of Product Use
1	AF
2	DE
3	FA
4	DA
5	EA
6	DF
7	FD
8	CD

Table M (cont.)

9	CF
10	CE
11	FC
12	AE
13	FE
14	DC
15	EF
16	ED
17	AB
18	BF
19	EB
20	EC
21	BA
22	BD
23	FB
24	BE
25	BC
26	AD
27	CB
28	DB
29	AC
30	CA

Basic design: $t = 7$, $b = 21$, $r = 6$, $\lambda = 1$
$q = 2$, $p = 15$

$pbq = 15(21)2 = 630$ panelists

$prq = 15(6)2 = 180$ judgments per sample

$p\lambda q = 15(1)2 = 30$ judgments per pair

Design: Repeat design in multiples of $bq = 42$

Panelist	Order of Product Use
1	EG
2	CD
3	FE
4	CB
5	BG
6	EF
7	GC
8	GD
9	EA
10	DB
11	BA
12	FA
13	FD

Table M (cont.)

14	GE
15	BC
16	DC
17	GA
18	FG
19	EB
20	CA
21	FC
22	DG
23	AE
24	AD
25	AC
26	BF
27	GF
28	CG
29	DA
30	BD
31	DE
32	ED
33	EC
34	DF
35	BE
36	AG
37	CE
38	AB
39	CF
40	GB
41	FB
42	AF

Basic design: $t = 8$, $b = 28$, $r = 7$, $\lambda = 1$
$q = 2$, $p = 15$

$pbq = 15(28)2 = 840$ panelists

$prq = 15(7)2 = 210$ judgments per sample

$p\lambda q = 15(1)2 = 30$ judgments per pair

Design: Repeat design in multiples of $bq = 56$

Panelist	Order of Product Use
1	GA
2	DB
3	AE
4	BA
5	DF
6	EH

Table M (cont.)

7	HF
8	GE
9	CG
10	HG
11	AB
12	BF
13	CH
14	AD
15	FD
16	AF
17	FB
18	DE
19	CD
20	BE
21	CA
22	BC
23	BD
24	BG
25	FA
26	GD
27	DA
28	GB
29	CB
30	DC
31	FH
32	HE
33	EC
34	HA
35	HB
36	GC
37	AG
38	ED
39	AH
40	EF
41	FC
42	HD
43	EB
44	FG

Table M (cont.)

45	CF
46	DG
47	EG
48	BH
49	HC
50	CE
51	FE
52	AC
53	GF
54	EA
55	DH
56	GH

Table N
Some popular Taguchi orthogonal array designs

First column pertains to the number of treatment conditions, runs or formulations. The remaining columns pertain to controllable factors.

L_4 Design: $L_4(2^3)$

Treatment	A	B	C
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1

Note: For two factors A and B, the C column becomes the A × B interaction. For three factors, the A × B interaction is confounded with factor C.

L_8 Design: $L_8(2^7)$

Treatment	A	B	C	D	E	F	G
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

L_{12} Design: $L_{12}(2^{11})$

Treatment	A	B	C	D	E	F	G	H	I	J	K
1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	2	2	2	2	2	2
3	1	1	2	2	2	1	1	1	2	2	2
4	1	2	1	2	2	1	2	2	1	1	2
5	1	2	2	1	2	2	1	2	1	2	1
6	1	2	2	2	1	2	2	1	2	1	1
7	2	1	2	2	1	1	2	2	1	2	1
8	2	1	2	1	2	2	2	1	1	1	2
9	2	1	1	2	2	2	1	2	2	1	1
10	2	2	2	1	1	1	1	2	2	1	2
11	2	2	1	2	1	2	1	1	1	2	2
12	2	2	1	1	2	1	2	1	2	2	1

Note: This design is a specially designed array, in that interactions are distributed more or less uniformly to all columns. It should not be used to analyze interactions. It is useful in optimization studies (See Chapter 8, Section 8.10).

Table N (cont.)

L_9 Design: $L_9(3^4)$

Treatment	A	B	C	D
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	1
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

Note: 1 = low level, 2 = medium level, 3 = high level

Source: Taguchi, G. and Konishi, S. 1987. Taguchi Methods: Orthogonal Arrays and Linear Graphs. American Supplier Institute, Inc., Dearborn, Michigan. Reproduced with permission of the American Supplier Institute.

REFERENCES

American Society of Testing Materials (ASTM). 1979. *ASTM Manual on Consumer Sensory Evaluation*. E.E. Schaefer, Ed., ASTM STP 682, Philadelphia, PA.

American Society of Testing Materials (ASTM). 1989. *Product Testing with Consumers for Research Guidance*. L.S. Wu, Ed., ASTM STP 1035, Philadelphia, PA.

Amerine, M.A., Pangborn, R.M., and Roessler, E.B. 1965. *Principles of Sensory Evaluation of Food*. Academic Press, New York.

Amerine, M.A. and Roessler, E.B. 1976. *Wines: Their Sensory Evaluation*. W.H. Freeman and Co., San Francisco, CA.

Aust, L.A., Gacula, Jr., M.C., Beard, S.A., and Washam II, R.W. 1985. Degree of difference test method in sensory evaluation. J. Food Sci. *50*:511–513.

Baird, J.C. and Noma, E. 1978. *Fundamentals of Scaling and Psychophysics*. John Wiley, New York.

Balaam, L.N. 1968. A two-period design with t × t experimental units. Biometrics *24*:61–73.

Barr, D.R. 1969. Using confidence intervals to test hypotheses. J. Quality Technol. *1*:256–258.

Basson, R.P. 1959. Incomplete block designs augmented with a repeated control. M.S. Thesis, Iowa State University, Ames, Iowa.

Bauer, D.F. 1978. Circular triads when not all paired comparisons are made. Biometrics *34*:458–461.

Blackwelder, W.C. 1982. Proving the null hypothesis in clinical trials. Controlled Clinical Trials *3*:345–353.

Blackwelder, W.C. and Chang, M.A. 1984. Sample size graphs for proving the null hypothesis. Controlled Clinical Trials *5*:97–105.

Box, G.E.P. 1988. Signal-to-noise, performance criteria, and transformation. Technometrics *30*:1–40.

Box, G.E.P. and Behnken, D.W. 1960. Some new three level designs for the study of quantitative variables. Technometrics *2*:455–475.

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. 1957. Multifactor experimental designs for exploring response surfaces. Ann. Math. Stat. *28*:195–241.

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

Box, G.E.P. and Wilson, K.B. 1951. On the experimental attainment of optimum conditions. J. Royal Stat. Soc., Ser. B *13*(1):1–38.

Bradley, R.A. and Terry, M.E. 1952. Rank analysis of incomplete block designs: I. The method of paired comparisons. Biometrika *39*:324.

Brown, B.W., Jr. 1980. The crossover experiment for clinical trials. Biometrics *36*:69.

Bullington, K.E., Hool, J.N., and Maghsoodloo, S. 1990. A simple method for obtaining resolution IV designs for use with Taguchi's orthogonal arrays. J. Quality Technol. *22*(4):260–264.

Cade, A.R. 1951. A method for testing degerming efficiency of hexochlorophene soaps. J. Soc. Cosmetic Chemists *2*:281-290.

Chambers IV, E., Bowers, J.A., and Dayton, A.D. 1981. Statistical designs and panel training/experience for sensory analysis. J. Food Sci. *46*:1902.

Chan, L.K., Cheng, S.W., and Spiring, F.A. 1991. A new measure of process capability: Cpm. J. Quality Technol. *20*:162-175.

Chen, T. 1960. Multiple Comparisons of Population Means. M.S. Thesis, Iowa State University, Ames, Iowa.

Claringbold, P.J. 1955. Use of the simplex design in the study of the joint action of related hormones. Biometrics *11*:174-185.

Cliff, M. and Wild, J. 1990. Development of SAS computer programs for sensory graphics. J. Sensory Studies *4*:249-260.

Cochran, W.G. 1983. *Planning and Analysis of Observational Studies*. John Wiley, New York.

Cochran, W.G. and Cox, G.M. 1957. *Experimental Designs*. John Wiley, New York.

Cornell, J.A. 1974. An alternative estimator for the recovery of interblock information when using incomplete block design. J. Food Sci. *39*:1254.

Cornell, J.A. 1981. *Experiments With Mixtures*. John Wiley, New York.

Cornell, J.A. 1983. *How to Run Mixture Experiments for Product Quality*. American Soc. Quality Control, Vol. 5, Milwaukee, Wisconsin.

Cornell, J.A. 1984. *How to Apply Response Surface Methodology*. American Soc. Quality Control, Vol. 8, Milwaukee, Wisconsin.

Council of Better Business Bureaus, Inc. 1990. PPG Industries, Inc. Lucite House and Trim Paint. NAD Case Report 20(9):31.

Crosby, P.B. 1979. *Quality is Free*. ASQC Quality Press, Milwaukee, WI.

Crowe, D.M., Willard, M.S., Murahata, R.I., Aust, L.B., and Gacula, Jr., M.C. 1988. Objective measurement of surfactant irritation by fiber optic spectroscopy. Contact Dermatitis *19*:192-194.

David, H.A. 1963. *The Method of Paired Comparisons*. Charles Griffin & Co., London.

Davies, O.L. and Hay, W.A. 1950. The construction and uses of fractional factorial designs in industrial research. Biometrics *6*:233.

Diptee, R., Smith, J.P., Alli, I., and Khanizadeh, S. 1989. Application of response surface methodology in protein extraction studies from brewer's spent grain. J. Food Processing Preservation *13*(6):457-474.

Dixon, W.J. 1950. Analysis of extreme values. Annals Math. Stat. *21*:488-506.

Dixon, W.J. 1951. Ratios involving extreme values. Annals Math. Stat. *22*:68-78.

Dixon, W.J. 1953. Processing data for outliers. Biometrics *9*:74-89.

Dixon, W.J. and Massey, F.J. 1957. *Introduction to Statistical Analysis*. McGraw-Hill, New York.

Dixon, W.J. and Mood, A.M. 1946. The statistical sign test. J. Amer. Stat. Assoc. *41*:557.

Draper, N.R. and Smith, H. 1981. *Applied Regression Analysis*. John Wiley, New York.

Duncan, D.B. 1955. Multiple range and multiple F tests. Biometrics *11*:1-42.

Dunn, O.J. 1964. Multiple comparisons using rank sums. Technometrics *6*:241-252.

Dunnett, C.W. and Gent, M. 1977. Significance testing to establish equivalence between treatments, with special reference to data in the form 2 × 2 tables. Biometrics *35*:593-692.

Dunn-Rankin, P. 1965. The True Probability Distribution of the Range of Rank Totals and its Application to Psychological Scaling. Ph.D. Thesis, Florida State University.

Dunn-Rankin, P. 1983. *Scaling Methods*. Lawrence Erlbaum Associates, Publisher, Hillside, NJ.

Durbin, J. 1951. Incomplete blocks in ranking experiments. Br. J. Psychol. (Part II) *4*:85–90.

Edwards, A.L. 1952. The scaling of the stimuli by the method of successive intervals. J. Appl. Psychol. *36*:118–122.

Edwards, A.L. 1957. *Techniques of Attitude Scale Construction*. Irvington Publishers, Inc., New York.

Federer, W.T. 1955. *Experimental Design*. Macmillan Co., New York.

Feigenbaum, A.V. 1991. *Total Quality Control*. ASQC Quality Press, Milwaukee, WI.

Ferber, R. (Ed.). 1974. *Handbook of Marketing Research*. McGraw-Hill, New York.

Finkey, M.B., Gacula, Jr., M.C., and Miller, J.K. 1988. Site to site variation in the soap chamber test. J. Sensory Studies *2*:293–300.

Finney, D.J. 1945. The fractional replication of factorial arrangements. Ann. Eugenics *12*:291–301.

Finney, D.J. 1946. Recent developments in the design of field experiments. III. Fractional replications. J. Agric. Sci. *36*:184–191.

Fisher, R.A. 1960. *The Design of Experiments*. 7th ed., Hafner Pub., New York.

Fishken, D. 1983. Consumer-oriented product optimization. Food Technol. *37*(11):49–52.

Freiman, J.A., Chalmers, T.C., Smith, H. and Kuebler, R.R. 1978. The importance of beta, type II error, and sample size in the design and interpretation of the randomized controlled trial: Survey of 71 negative trials. N. Engl. J. Med. *299*:690–694.

Freund, R.J. and Littell, R.C. 1986. *SAS System for Regression*. SAS Institute Inc., Cary, NC.

Gacula, Jr., M.C. 1978. Analysis of incomplete block designs with reference samples in every block. J. Food Sci. *43*:1461–1466.

Gacula, Jr., M.C. 1987. Some issues in the design and analysis of sensory data: Revisited. J. Sensory Studies *2*:169–185.

Gacula, Jr., M.C. and Kubala, J.J. 1972. Data analysis: Interblock and intrablock estimates of variance on taste panel data. J. Food Sci. *37*:832–836.

Gacula, Jr., M.C., Reaume, J.B., Morgan, K.J., and Luckett, R.L. 1971. The statistical aspect of the correlation between objective and subjective measurements. J. Food Sci. *36*:185–189.

Gacula, Jr., M.C., Rutenbeck, S.K., Campbell, J.F., Giovanni, M.E., Gardze, C.A., and Washam II, R.W. 1986. Some sources of bias in consumer testing. J. Sensory Studies *1*:175–182.

Gacula, Jr., M.C. and Singh, J. 1984. *Statistical Methods in Food and Consumer Research*. Academic Press, Orlando, FL.

Gacula, Jr., M.C., Washam II, W.W., Beard, S.A., and Heinze, J.E. 1986. Estimates of carry-over effects in two-product home usage consumer tests. J. Sensory Studies *1*:47–53.

Galvez, F.C.P., Resurreccion, A.V.A, and Koehler, P.E. 1990. Optimization of processing peanut beverage. J. Sensory Studies *5*:1–17.

Gart, J.J. 1969. An exact test for comparing matched proportions in crossover designs. Biometrika *56*:75–80.

Giovanni, M.E. 1983. Response surface methodology and product optimization. Food Technol. *37*(11):41–44, 83.

Giovanni, M.E. and Pangborn, R.M. 1983. Measurement of taste intensity and degree of liking of beverages by graphic rating scale and magnitude estimation. J. Food Sci. *48*:1175–1181.

Gorden, R.L. 1977. *Unidimensional Scaling of Social Variables*. The Free Press, New York.
Gorman, J.W. and Hinman, J.E. 1962. Simplex lattice designs for multicomponent systems. Technometrics *4*:463–487.
Green, J.R. 1971. Testing departure from a regression, without using replication. Technometrics *13*:609–615.
Gridgeman, N.T. 1963. Significance and adjustment in paired comparisons. Biometrics *19*:213–228.
Griffin, R. and Stauffer, L. 1991. Product optimization in central-location testing and subsequent validation and calibration in home-use testing. J. Sensory Studies *4*:231–240.
Grizzle, J.E. 1965. The two-period changeover design and its use in clinical trials. Biometrics *21*:467–480.
Grubbs, F.E. 1950. Sample size criteria for testing outlying observations. Annals Math. Stat. *21*:27–58.
Grubbs, F.E. and Beck, G. 1972. Extension of sample sizes and percentage points for significance tests of outlying observations. Technometrics *14*:847–854.
Guilford, J.P. 1954. *Psychometric Methods*. McGraw-Hill, New York.
Hanson, H.L., Kline, L., and Lineweaver, H. 1951. Application of balanced incomplete block design to scoring of ten dried egg samples. Food Technol. *5*:9–13.
Hare, L.B. 1974. Mixture designs applied to food formulation. Food Technology *28*(3):50–62.
Hartley, H.O. 1959. Smallest composite designs for quadratic response surfaces. Biometrics *15*:611–624.
Hills, M. and Armitage, P. 1979. The two-period cross-over clinical trial. Br. J. Clin. Pharmac. *8*:7–20.
Horst, P. 1932. A method for determining the absolute affective value of a series of stimulus situations. J. Educ. Psychol. *23*:418–440.
Jao, Y.C., Larkin, J.W., Chen, A.H. and Goldstein, W.E. 1982. Response surface analysis of egg white gelling properties in a meat loaf analog. J. Food Proc. Eng. *4*:227–239.
Joanes, D.N. 1985. On a rank sum test due to Kramer. J. Food Sci. *50*:1442–1444.
Joglekar, G., Schuenemeyer, J.H and LaRiccia, V. 1989. Lack of fit testing when replicates are not available. Amer. Statistician *43*:135–143.
John, P.W.M. 1971. *Statistical Design and Analysis of Experiments*. Macmillan, New York.
John, P.W.M. 1980. *Incomplete Block Designs*. Marcel Dekker, New York.
Jones, B. and Kenward, M.G. 1989. *Design and Analysis of Cross-Over Trials*. Chapman and Hall, London.
Jones, D.A. and Karson, M.J. 1972. On the use of the confidence regions to test hypotheses. J. Quality Technol. *4*:156–158.
Jones, F.N. 1974. History of psychophysics and judgement. In *Handbook of Perception, Vol. II* (Carterette, E.C. and Friedman, M.P., eds.). Academic Press, New York.
Juran, J.M. (Ed.). 1980. *Quality Control Handbook*. McGraw Hill, New York.
Kacker, R.N., Lagergren, E.S. and Filliben, J.J. 1991. Taguchi's fixed-element arrays are fractional factorials. J. Quality Technol. *23*:107–116.
Kacker, R.N. and Tsui, K. 1990. Interaction graphs: Graphical aids for planning experiments. J. Quality Technol. *22*:1–14.
Kempthorne, O. 1952. *The Design and Analysis of Experiments*. John Wiley, New York.
Kempthorne, O. 1947. A simple approach to confounding and fractional replication in factorial experiments. Biometrika *34*:255–272.

Kershner, R.P. and Federer, W.T. 1981. Two-treatment crossover designs for estimating a variety of effects. J. Amer. Stat. Assoc. *76*:612–618.

Kirk, R.E. 1968. *Experimental Design: Procedures for the Behavioral Sciences*. Brooks/Cole, Belmont, CA.

Koch, G.G. 1972. The use of nonparametric methods in the statistical analysis of the two-period change-over design. Biometrics *28*:577–584.

Kramer, A. 1963. Revised tables for determining significance of differences. Food Technol. *17*:124–125.

Kramer, A., Kahan, G., Cooper, D., and Papavasiliou, A. 1974. A nonparametric ranking method for the statistical evaluation of sensory data. Chemical Senses and Flavor *1*:121.

Kurotori, I.S. 1966. Experiments with mixtures of components having lower bounds. Industrial Quality Control *22*:592–596.

Kvalseth, T.O. 1985. Cautionary note about R^2. Amer. Statistician *39*:279–285.

Lagrange, V. and Norback, J.P. 1987. Product optimization and the acceptor set size. J. Sensory Studies *2*:119–136.

Larmond, E. 1967. Methods for Sensory Evaluation of Food. Publ. Canada Dept. Agric. No. 1284.

Laska, E.M., Meisner, M. and Kushner, H.B. 1983. Optimal crossover designs in the presence of carryover effects. Biometrics *39*:1089–1091.

Lawless, H.T. and Malone, G.T. 1986a. The discriminative efficiency of common scaling methods. J. Sensory Studies *1*:85–98.

Lawless, H.T. and Malone, G.T. 1986b. A comparison of rating scales: Sensitivity, replicates, and measurement. J. Sensory Studies *1*:155–174.

Lehmann, E.L. 1986. *Testing Statistical Hypotheses*. John Wiley, New York.

Lundahl, D.S. and McDaniel, M. 1988. The panelist effect-fixed or random? J. Sensory Studies *3*:113–121.

Maghsoodloo, S. 1990. The exact relation of Taguchi's signal-to-noise ratio to his quality loss function. J. Quality Technol. *22*(1):57–67.

Marquardt, D.W. and Snee, R.D. 1974. Test statistics for mixture models. Technometrics *16*:533–537.

McBride, R.L. 1986. Hedonic rating of food: single or side-by-side sample presentation? J. Food Technol. *21*:355–363.

McDaniel, M.R. and Sawyer, F.M. 1981a. Descriptive analysis of whiskey sour formulations: Magnitude estimation versus a 9-point category scale. J. Food Sci. *46*:178–181.

McDaniel, M.R. and Sawyer, F.M. 1981b. Preference testing of whiskey sour formulations: Magnitude estimation versus a 9-point category scale. J. Food Sci. *46*:182–185.

McLean, R.A. and Anderson, V.L. 1966. Extreme vertices design of mixture experiments. Technometrics *8*:447–456.

McLean, R.A. and Anderson, V.L. 1984. *Applied Factorial and Fractional Designs*. Marcel Dekker, New York.

McLellan, M.R., Barnard, J., and Queale, D.T. 1984. Sensory analysis of carbonated apple juice using response surface methodology. J. Food Sci. *49*:1595–1597.

Meilgaard, M., Civille, G.V., and Carr, B.T. 1987. *Sensory Evaluation Techniques*. CRC Press, Boca Raton, FL.

Metzler, C.M. 1974. Bioavailability — a problem in equivalence. Biometrics *30*:309–317.

Microsoft Corporation. 1992. Microsoft Draw. Microsoft Corp., Redmond, WA.

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

Montgomery, D.C. 1991b. *Introduction to Statistical Quality Control*. John Wiley, New York.

Moskowitz, H.R. 1977. Magnitude estimation: Notes on what, how, when and why to use it. J. Food Quality *1*:195-228.

Moskowitz, H.W. 1983. *Product Testing and Sensory Evaluation of Foods*. Food & Nutrition Press, Trumbull, CT.

Moskowitz, H.W. 1984. *Cosmetic Product Testing*. Marcel Dekker, New York.

Moskowitz, H.R. (Ed.). 1988. *Applied Sensory Analysis of Foods. Vols. I and II*. CRC Press, Boca Raton, FL.

Moskowitz, H.R. and Sidel, J. 1971. Magnitude and hedonic scales of food acceptability. J. Food Sci. *36*:677-680.

Mosteller, F. 1951a. Remarks on the method of paired comparisons: II. The effect of an aberrant standard deviation when equal standard deviations and equal correlations are assumed. Psychometrika *16*:203-206.

Mosteller, F. 1951b. Remarks on the method of paired comparisons: III. A test of significance for paired comparison when equal standard deviations and equal correlations are assumed. Psychometrika *16*:207-218.

Mudahar, G.S., Toledo, R.T., and Jen, J.J. 1990. A response surface methodology approach to optimize potato dehydration process. J. Food Processing Preservation *14*(2):93.

Mullen, K. and Ennis, D. 1985. Fractional factorials in product development. Food Technol. *39*(5):90-103.

Myers, R.M. 1971. *Response Surface Methodology*. Allyn and Bacon, Boston, MA.

Nair, K.R. 1952. Tables of percentage points of the studentized extreme deviate from the sample mean. Biometrika *31*:189-191.

Nally, C.L. 1987. Implementation of consumer taste panels. J. Sensory Studies *2*:77-84.

Nam, J. 1971. On two tests for comparing matched proportions. Biometrics *27*:945-959.

Natrella, M.G. 1960. The relation between confidence intervals and tests of significance. Amer. Statistician *14*:20-22,38.

Nemenyi, P. 1963. Distribution-Free Multiple Comparisons. Ph.D. Thesis, Princeton University.

Neter, J. and Wasserman, W. 1974. *Applied Linear Statistical Models*. Irwin, Homewood, IL.

O'Mahony, M. 1986. *Sensory Evaluation of Food: Statistical Methods and Procedures*. Marcel Dekker, New York.

Parducci, A. 1982. Category ratings: Still more contextual effects. In *Social Attitudes and Psychological Measurement*, (B. Wegener, ed.) Erlbaum, Hillside, NJ.

Pearce, S.C. 1960. Supplemented balance. Biometrika *47*:263-271.

Pearce, J.H., Korth, B. and Warren, C.B. 1986. Evaluation of three scaling methods for hedonics. J. Sensory Studies *1*:27-46.

Peng, K.C. 1967. *The Design and Analysis of Scientific Experiments*. Addison-Wesley, MA.

Peryam, D.R. and Girardot, H.F. 1952. Advanced taste-test method. Food Eng. *24*:58-61,194.

Piepel, G.F. and Cornell, J.A. 1987. Design for mixture-amount experiments. J. Quality Technol. *19*:11-28.

Piggott, J.R. (Ed.). 1986. *Statistical Procedures in Food Research*. Elsevier Applied Science, London.

Plackett, R.L. and Burman, J.P. 1946. The design of optimum multifactorial experiments. Biometrika *33*:305-325.

Quenouille, M.H. 1965. *Fundamentals of Statistical Reasoning*. Charles Griffin & Co., London.

Rao, S.S. 1984. *Optimization: Theory and Applications*. John Wiley, New York.

Riskey, D.R. 1986. Use and abuses of category scales in sensory measurement. J. Sensory Studies *1*:217-236.

Ryan, T.P. 1989. *Statistical Methods for Quality Improvement*. John Wiley, New York.

Saxena, S.K. and Nigam, A.K. 1977. Restricted exploration of mixtures by symmetric-simplex designs. Technometrics *19*:47-52.

Scheffe, H. 1952. An analysis of variance for paired comparisons. J. Amer. Stat. Assoc. *47*:381-400.

Scheffe, H. 1958. Experiments with mixtures. J. Royal Stat. Soc., Ser. B. *20*:344-360.

Scheffe, H. 1963. The simplex-centroid design for experiments with mixtures. J. Royal Stat. Soc., Ser. B, *25*:235-263.

Schilling, E.G. 1982. *Acceptance Sampling in Quality Control*. Marcel Dekker, New York.

Schutz, H.G. 1983. Multiple regression approach to optimization. Food Technol. *37*(11):46-48,62.

Seshadri, V. 1963. Combining unbiased estimators. Biometrics *19*:163-170.

Shand, P.J., Hawrysh, Z.J., Hardin, R.T., and Jermiah, L.E. 1985. Descriptive sensory assessment of beef steaks by category scaling, line scaling, and magnitude estimation. J. Food Sci. *50*:495-500.

Shiffler, R.E. 1988. Maximum Z scores and outliers. Amer. Statistician *42*:79-80.

Shillington, E.R. 1979. Testing lack of fit without replication. Canadian J. Statistics *7*:137-146.

Shirley, E. 1976. The use of confidence intervals in biopharmaceutics. J. Pharm. Pharmacol. *28*:312-313.

Sidel, J. and Stone, H. 1983. An introduction to optimization research. Food Technol. *37*(11):36-38.

Sincich, T. 1986. *Business Statistics by Example*. Dellen Publishing Co., San Francisco.

Skillings, J.H. and Mack, G.A. 1981. On the use of a Friedman-type statistic in balanced and unbalanced block designs. Technometrics *23*:171-177.

Slater, P. 1961. Inconsistencies in a schedule of paired comparisons. Biometrika *48*:303-312.

Snedecor, G.W. and Cochran, W.G. 1967. *Statistical Methods*. The Iowa State University Press, Ames, Iowa.

Snee, R.D. 1971. Design and analysis of mixture experiments. J. Quality Technol. *3*:159-169.

Snee, R.D. and Marquardt, D.W. 1976. Extreme vertices designs for linear mixture models. Technometrics *16*:399-408.

Spriet, A. and Beiler, D. 1979. When can nonsignificantly different treatments be considered as equivalent? Brit. J. Clin. Pharmacol. *7*:623-624.

Statistical Engr. Lab. 1957. Fractional factorial experiment designs for factors at two levels. National Bu. Stand. Appl. Math. Series 48.

Steel, R.G.D. 1961. Some rank sum multiple comparison tests. Biometrics *17*:539-552.

Steel, R.G.D. and Torrie, J.H. 1960. *Principles and Procedures of Statistics*. McGraw-Hill, New York.

Stevens, S.S. 1946. On the theory of scales of measurement. Science *103*(2684):677-680.

Stevens, S.S. 1951. Mathematics, measurement, and psychophysics. In *Handbook of Experimental Psychology*, (S.S. Stevens, ed.), p.1-49. John Wiley, New York.

Stevens, S.S. 1961. To honor Fechner and repeal his law. Science *133*:80-86.

Stevens, S.S. and Galanter, E.H. 1957. Ratio scales and category scales for a dozen perceptual continua. J. Exp. Psychol. *54*:377–411.

Stone, H. and Sidel, J. 1985. *Sensory Evaluation Practices*. Academic Press, Orlando, FL.

Stone, H., Sidel, J., Oliver, S., Woolsey, A. and Singleton, R.C. 1974. Sensory evaluation by quantitative descriptive analysis. Food Technol. *28*(11):24–34.

Taguchi, G. 1985. A tutorial on quality control and assurance: The Taguchi Methods. ASA Annual Meeting, Las Vegas, Nevada.

Taguchi, G. 1986. Introduction to Quality Engineering. Asian Productivity Organization, Tokyo, Japan.

Taguchi, G. and Konishi, S. 1987. Orthogonal Arrays and Linear Graphs. American Supplier Institute, Inc., Dearborn, Michigan.

Taguchi, G. and Wu, Y. 1979. Introduction to Off-line Quality Control. Central Japan Quality Control Assoc., Tokyo, Japan.

Thurstone, L.L. 1927. Psychophysical analysis. Amer. J. Psychol. *38*:369–389.

Thurstone, L.L. 1959. *The Measurement of Values*. The University of Chicago Press, Chicago, IL.

Thurstone, L.L. and Jones, L.V. 1957. The rational origin for measuring subjective values. J. Amer. Stat. Assoc. *52*:458–471.

Tobach, E., Smith, M. Rose, G., and Richter, D. 1967. A table for making rank sum multiple paired comparisons. Technometrics *9*:561–567.

Torgerson, W.S. 1958. *Theory and Methods of Scaling*. John Wiley, New York.

Westlake, W.J. 1965. Composite designs based on irregular fractions of factorials. Biometrics *21*:324–336.

Westlake, W.J. 1972. Use of confidence intervals in analysis of comparative bioavailability trials. J. Pharmaceutical Sci. *61*:1340–1341.

Westlake, W.J. 1976. Symmetrical confidence intervals for bioequivalence trials. Biometrics *32*:741–744.

Wyckham, R.G. 1987. Implied superiority claims. J. Advertising Res. *27*:54–63 (Feb./March).

Yates, F. 1936. Incomplete randomized blocks. Ann. Eugenics *7*:121–140.

Yates, F. 1937. The design and analysis of factorial experiments. Tech. Commun. Bur. Soil Sci. No. 35, Harpenden, England.

Yates, F. 1940. The recovery of interblock information in balanced incomplete block design. Ann. Eugenics *10*:317–325.

Younger, M.S. 1979. *A Handbook for Linear Regression*. Duxbury Press, North Scituate, MA.

Wilcoxon, F. and Wilcox, R.A. 1964. Some Rapid Approximate Statistical Procedures. Lederle Laboratories, Pearl River, New York.

Williams, E.R. and Jones, P.N. 1979. Balanced incomplete block designs with a reference sample in each block. J. Food Sci. *44*:1790–1791.

Zelek, E.F. 1990. Legal aspects of sensory analysis. Food Technol. *44*(11):168–174.

INDEX