PETER GOOS
BRADLEY JONES

OPTIMAL DESIGN OF EXPERIMENTS

A Case Study Approach

WILEY

Optimal Design of Experiments

A Case Study Approach

Optimal Design of Experiments

A Case Study Approach

Peter Goos
University of Antwerp and Erasmus University Rotterdam

Bradley Jones
JMP Division of SAS

A John Wiley & Sons, Ltd., Publication

This edition first published 2011

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

Library of Congress Cataloging-in-Publication Data

Goos, Peter.
Optimal design of experiments: a case study approach / Peter Goos and Bradley Jones.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-74461-1 (hardback)
1. Industrial engineering–Experiments–Computer-aided design. 2. Experimental design–Data processing. 3. Industrial engineering–Case studies. I. Jones, Bradley. II. Title.
T57.5.G66 2011
670.285–dc22

2011008381

A catalogue record for this book is available from the British Library.

Print ISBN: 978-0-470-74461-1
ePDF ISBN: 978-1-119-97400-0
oBook ISBN: 978-1-119-97401-7
ePub ISBN: 978-1-119-97616-5
Mobi ISBN: 978-1-119-97617-2

Set in 10/12pt Times by Aptara Inc., New Delhi, India.

To Marijke, Bas, and Loes

To Roselinde

Contents

Preface

Design of experiments is a powerful tool for understanding systems and processes. In practice, this understanding often leads immediately to improvements. We present optimal design of experiments as a general and flexible method for applying design of experiments. Our view is that optimal design of experiments is an appropriate tool in virtually any situation that suggests the possible use of design of experiments.

Books on application areas in statistics or applied mathematics, such as design of experiments, can present daunting obstacles to the nonexpert. We wanted to write a book on the practical application of design of experiments that would appeal to new practitioners and experts alike. This is clearly an ambitious goal and we have addressed it by writing a different kind of book.

Each chapter of the book contains a case study. The presentation of the case study is in the form of a play where two consultants, Brad and Peter, of the (fictitious) Intrepid Stats consulting firm, help clients in various industries solve practical problems. We chose this style to make the presentation of the core concepts of each chapter both informal and accessible.

This style is by no means unique. The use of dialogs dates all the way back to the Greek philosopher Plato. More recently, Galileo made use of this style to introduce scientific ideas. His three characters were: the teacher, the experienced student, and the novice.

Though our case studies involve scripted consulting sessions, we advise readers not to copy our consulting style when collaborating on their own design problems. In the interest of a compact exposition of the key points of each case, we skip much of the necessary information gathering involved in competent statistical consulting and problem solving.

We chose our case studies to show just how general and flexible the optimal design of experiments approach is. We start off by a chapter dealing with a simple comparative experiment. The next two chapters deal with a screening experiment and a follow-up experiment in a biotechnology firm. In Chapter 4, we show how a designed response surface experiment contributes to the development of a robust production process in food packaging. In Chapter 5, we set up a response surface experiment to maximize the yield of a chemical extraction process. Chapter 6 deals with an experiment, similar in structure to mixture experiments in the chemical and pharmaceutical industries, aimed at improving the finishing of aluminum sheets. In Chapters 7 and 8, we apply the optimal design of experiments approach to a vitamin

stability experiment and a pastry dough experiment run over different days, and we demonstrate that this offers protection against day-to-day variation in the outcomes. In Chapter 9, we show how to take into account a priori information about the experimental units and how to deal with a time trend in the experimental results. In Chapter 10, we set up a wind tunnel experiment that involves factors whose levels are hard to change. Finally, in Chapter 11, we discuss the design of a battery cell experiment spanning two production steps.

Because our presentation of the case studies is often light on mathematical and statistical detail, each chapter also has a section that we call a "Peek into the black box." In these sections, we provide a more rigorous underpinning for the various techniques we employ in our case studies. The reader may find that there is not as much material in these sections on data analysis as might be expected. Many books on design of experiments are mostly about data analysis rather than design generation, evaluation, and comparison. We focus much of our attention in these peeks into the black box on explaining what the reader can anticipate from the analysis, before actually acquiring the response data. In nearly every chapter, we have also included separate frames, which we call "Attachments," to discuss topics that deserve special attention.

We hope that our book will appeal to the new practitioner as well as providing some utility to the expert. Our fondest wish is to empower more experimentation by more people. In the words of Cole Porter, "Experiment and you'll see!"

Acknowledgments

We would like to express our gratitude to numerous people who helped us in the process of writing this book.

First, we would like to thank Chris Nachtsheim for allowing us to use the scenario for the "mixture" experiment in Chapter 6, and Steven Gilmour for providing us with details about the pastry dough experiment in Chapter 7. We are also grateful to Ives Bourgonjon, Ludwig Claeys, Pascal Dekoninck, Tim De Rydt, Karen Dewilde, Heidi Dufait, Toine Machiels, Carlo Mol, and Marc Pauwels whose polypropylene project in Belgium, sponsored by Flanders' Drive, provided inspiration for the case study in Chapter 9.

A screening experiment described in Bie et al. (2005) provided inspiration for the case study in Chapters 2 and 3, while the work of Brenneman and Myers (2003) stimulated us to work out the response surface study involving a categorical factor in Chapter 4. We adapted the case study involving a constrained experimental region in Chapter 5 from an example in Box and Draper (1987). The vitamin stability experiment in Loukas (1997) formed the basis of the blocked screening experiment in Chapter 8. We turned the wind tunnel experiment described in Simpson et al. (2004) and the battery cell experiment studied in Vivacqua and Bisgaard (2004) into the case studies in Chapters 10 and 11.

Finally, we would like to thank Marjolein Crabbe, Marie Gaudard, Steven Gilmour, J. Stuart Hunter, Roselinde Kessels, Kalliopi Mylona, John Sall, Eric Schoen, Martina Vandebroek, and Arie Weeren for proofreading substantial portions of this book. Of course, all remaining errors are our own responsibility.

Heverlee,
Peter Goos

Cary,
Bradley Jones

January 2011

1

A simple comparative experiment

1.1 Key concepts

1. Good experimental designs allow for precise estimation of one or more unknown quantities of interest. An example of such a quantity, or parameter, is the difference in the means of two treatments. One parameter estimate is more precise than another if it has a smaller variance.
2. Balanced designs are sometimes optimal, but this is not always the case.
3. If two design problems have different characteristics, they generally require the use of different designs.
4. The best way to allocate a new experimental test is at the treatment combination with the highest prediction variance. This may seem counterintuitive but it is an important principle.
5. The best allocation of experimental resources can depend on the relative cost of runs at one treatment combination versus the cost of runs at a different combination.

Is A different from B? Is A better than B? This chapter shows that doing the same number of tests on A and on B in a simple comparative experiment, while seemingly sensible, is not always the best thing to do. This chapter also defines what we mean by the best or optimal test plan.

Optimal Design of Experiments: A Case Study Approach, First Edition. Peter Goos and Bradley Jones.

1.2 The setup of a comparative experiment

Peter and Brad are drinking Belgian beer in the business lounge of Brussels Airport. They have plenty of time as their flight to the United States is severely delayed due to sudden heavy snowfall. Brad has just launched the idea of writing a textbook on tailor-made design of experiments.

[Brad] I have been playing with the idea for quite a while. My feeling is that design of experiments courses and textbooks overemphasize standard experimental plans such as full factorial designs, regular fractional factorial designs, other orthogonal designs, and central composite designs. More often than not, these designs are not feasible due to all kinds of practical considerations. Also, there are many situations where the standard designs are not the best choice.

[Peter] You don't need to convince me. What would you do instead of the classical approach?

[Brad] I would like to use a case-study approach. Every chapter could be built around one realistic experimental design problem. A key feature of most of the cases would be that none of the textbook designs yields satisfactory answers and that a flexible approach to design the experiment is required. I would then show that modern, computer-based experimental design techniques can handle real-world problems better than standard designs.

[Peter] So, you would attempt to promote optimal experimental design as a flexible approach that can solve any design of experiments problem.

[Brad] More or less.

[Peter] Do you think there is a market for that?

[Brad] I am convinced there is. It seems strange to me that, even in 2011, there aren't any books that show how to use optimal or computer-based experimental design to solve realistic problems without too much mathematics. I'd try to focus on how easy it is to generate those designs and on why they are often a better choice than standard designs.

[Peter] Do you have case studies in mind already?

[Brad] The robustness experiment done at Lone Star Snack Foods would be a good candidate. In that experiment, we had three quantitative experimental variables and one categorical. That is a typical example where the textbooks do not give very satisfying answers.

[Peter] Yes, that is an interesting case. Perhaps the pastry dough experiment is a good candidate as well. That was a case where a response surface design was run in blocks, and where it was not obvious how to use a central composite design.

[Brad] Right. I am sure we can find several other interesting case studies when we scan our list of recent consulting jobs.

[Peter] Certainly.

[Brad] Yesterday evening, I tried to come up with a good example for the introductory chapter of the book I have in mind.

[Peter] Did you find something interesting?

[Brad] I think so. My idea is to start with a simple example. An experiment to compare two population means. For example, to compare the average thickness of cables produced on two different machines.

[Peter] So, you'd go back to the simplest possible comparative experiment?

[Brad] Yep. I'd do so because it is a case where virtually everybody has a clear idea of what to do.

[Peter] Sure. The number of observations from the two machines should be equal.

[Brad] Right. But only if you assume that the variance of the thicknesses produced by the two machines is the same. If the variances of the two machines are different, then a 50–50 split of the total number of observations is no longer the best choice.

[Peter] That could do the job. Can you go into more detail about how you would work that example?

[Brad] Sure.

Brad grabs a pen and starts scribbling key words and formulas on his napkin while he lays out his intended approach.

[Brad] Here we go. We want to compare two means, say μ_1 and μ_2, and we have an experimental budget that allows for, say, $n = 12$ observations, n_1 observations from machine 1 and $n - n_1$ or n_2 observations from machine 2. The sample of n_1 observations from the first machine allows us to calculate a sample mean $\overline{X}_1$ for the first machine, with variance σ^2/n_1. In a similar fashion, we can calculate a sample mean $\overline{X}_2$ from the n_2 observations from the second machine. That second sample mean has variance σ^2/n_2.

[Peter] You're assuming that the variance in thickness is σ^2 for both machines, and that all the observations are statistically independent.

[Brad] Right. We are interested in comparing the two means, and we do so by calculating the difference between the two sample means, $\overline{X}_1 - \overline{X}_2$. Obviously, we want this estimate of the difference in means to be precise. So, we want its variance

$$\text{var}(\overline{X}_1 - \overline{X}_2) = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2\Big(\frac{1}{n_1} + \frac{1}{n_2}\Big)$$

or its standard deviation

$$\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}} = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

to be small.

[Peter] Didn't you say you would avoid mathematics as much as possible?

[Brad] Yes, I did. But we will have to show a formula here and there anyway. We can talk about this later. Stay with me for the time being.

Brad empties his Leffe, draws the waiter's attention to order another, and grabs his laptop.

[Brad] Now, we can enumerate all possible experiments and compute the variance and standard deviation of $\overline{X}_1 - \overline{X}_2$ for each of them.

Table 1.1 Variance of sample mean difference for different sample sizes n_1 and n_2 for $\sigma^2 = 1$.

n_1	n_2	$\text{var}(\overline{X}_1 - \overline{X}_2)$	$\sigma_{\overline{X}_1 - \overline{X}_2}$	Efficiency (%)
1	11	1.091	1.044	30.6
2	10	0.600	0.775	55.6
3	9	0.444	0.667	75.0
4	8	0.375	0.612	88.9
5	7	0.343	0.586	97.2
6	6	0.333	0.577	100.0
7	5	0.343	0.586	97.2
8	4	0.375	0.612	88.9
9	3	0.444	0.667	75.0
10	2	0.600	0.775	55.6
11	1	1.091	1.044	30.6

Before the waiter replaces Brad's empty glass with a full one, Brad has produced Table 1.1. The table shows the 11 possible ways in which the $n = 12$ observations can be divided over the two machines, and the resulting variances and standard deviations.

[Brad] Here we go. Note that I used a σ^2 value of one in my calculations. This exercise shows that taking n_1 and n_2 equal to six is the best choice, because it results in the smallest variance.

[Peter] That confirms traditional wisdom. It would be useful to point out that the σ^2 value you use does not change the choice of the design or the relative performance of the different design options.

[Brad] Right. If we change the value of σ^2, then the 11 variances will all be multiplied by the value of σ^2 and, so, their relative magnitudes will not be affected. Note that you don't lose much if you use a slightly unbalanced design. If one sample size is 5 and the other is 7, then the variance of our sample mean difference, $\overline{X}_1 - \overline{X}_2$, is only a little bit larger than for the balanced design. In the last column of the table, I computed the efficiency for the 11 designs. The design with sample sizes 5 and 7 has an efficiency of $0.333/0.343 = 97.2\%$. So, to calculate that efficiency, I divided the variance for the optimal design by the variance of the alternative.

[Peter] OK. I guess the next step is to convince the reader that the balanced design is not always the best choice.

Brad takes a swig of his new Leffe, and starts scribbling on his napkin again.

[Brad] Indeed. What I would do is drop the assumption that both machines have the same variance. If we denote the variances of machines 1 and 2 by σ_1^2 and σ_2^2, respectively, then the variances of $\overline{X}_1$ and $\overline{X}_2$ become σ_1^2/n_1 and σ_2^2/n_2. The variance of our sample mean difference $\overline{X}_1 - \overline{X}_2$ then is

$$\text{var}(\overline{X}_1 - \overline{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2},$$

Table 1.2 Variance of sample mean difference for different sample sizes n_1 and n_2 for $\sigma_1^2 = 1$ and $\sigma_2^2 = 9$.

n_1	n_2	$\text{var}(\overline{X}_1 - \overline{X}_2)$	$\sigma_{\overline{X}_1 - \overline{X}_2}$	Efficiency (%)
1	11	1.818	1.348	73.3
2	10	1.400	1.183	95.2
3	9	1.333	1.155	100.0
4	8	1.375	1.173	97.0
5	7	1.486	1.219	89.7
6	6	1.667	1.291	80.0
7	5	1.943	1.394	68.6
8	4	2.375	1.541	56.1
9	3	3.111	1.764	42.9
10	2	4.600	2.145	29.0
11	1	9.091	3.015	14.7

so that its standard deviation is

$$\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

[Peter] And now you will again enumerate the 11 design options?

[Brad] Yes, but first I need an a priori guess for the values of σ_1^2 and σ_2^2. Let's see what happens if σ_2^2 is nine times σ_1^2.

[Peter] Hm. A variance ratio of nine seems quite large.

[Brad] I know. I know. I just want to make sure that there is a noticeable effect on the design.

Brad pulls his laptop a bit closer and modifies his original table so that the thickness variances are $\sigma_1^2 = 1$ and $\sigma_2^2 = 9$. Soon, he produces Table 1.2.

[Brad] Here we are. This time, a design that requires three observations from machine 1 and nine observations from machine 2 is the optimal choice. The balanced design results in a variance of 1.667, which is 25% higher than the variance of 1.333 produced by the optimal design. The balanced design now is only $1.333/1.667 = 80\%$ efficient.

[Peter] That would be perfect if the variance ratio was really as large as nine. What happens if you choose a less extreme value for σ_2^2? Can you set σ_2^2 to 2?

[Brad] Sure.

A few seconds later, Brad has produced Table 1.3.

[Peter] This is much less spectacular, but it is still true that the optimal design is unbalanced. Note that the optimal design requires more observations from the machine with the higher variance than from the machine with the lower variance.

[Brad] Right. The larger value for n_2 compensates the large variance for machine 2 and ensures that the variance of $\overline{X}_2$ is not excessively large.

Table 1.3 Variance of sample mean difference for different sample sizes n_1 and n_2 for $\sigma_1^2 = 1$ and $\sigma_2^2 = 2$.

n_1	n_2	$\text{var}(\overline{X}_1 - \overline{X}_2)$	$\sigma_{\overline{X}_1 - \overline{X}_2}$	Efficiency (%)
1	11	1.182	1.087	41.1
2	10	0.700	0.837	69.4
3	9	0.556	0.745	87.4
4	8	0.500	0.707	97.1
5	7	0.486	0.697	100.0
6	6	0.500	0.707	97.1
7	5	0.543	0.737	89.5
8	4	0.625	0.791	77.7
9	3	0.778	0.882	62.4
10	2	1.100	1.049	44.2
11	1	2.091	1.446	23.2

[Peter, pointing to Table 1.3] Well, I agree that this is a nice illustration in that it shows that balanced designs are not always optimal, but the balanced design is more than 97% efficient in this case. So, you don't lose much by using the balanced design when the variance ratio is closer to 1.

Brad looks a bit crestfallen and takes a gulp of his beer while he thinks of a comeback line.

[Peter] It would be great to have an example where the balanced design didn't do so well. Have you considered different costs for observations from the two populations? In the case of thickness measurements, this makes no sense. But imagine that the two means you are comparing correspond to two medical treatments. Or treatments with two kinds of fertilizers. Suppose that an observation using the first treatment is more expensive than an observation with the second treatment.

[Brad] Yes. That reminds me of Eric Schoen's coffee cream experiment. He was able to do twice as many runs per week with one setup than with another. And he only had a fixed number of weeks to run his study. So, in terms of time, one run was twice as expensive as another.

[Peter, pulling Brad's laptop toward him] I remember that one. Let us see what happens. Suppose that an observation from population 1, or an observation with treatment 1, costs twice as much as an observation from population 2. To keep things simple, let the costs be 2 and 1, and let the total budget be 24. Then, we have 11 ways to spend the experimental budget I think. One extreme option takes one observation for treatment 1 and 22 observations for treatment 2. The other extreme is to take 11 observations for treatment 1 and 2 observations for treatment 2. Each of these extreme options uses up the entire budget of 24. And, obviously, there are a lot of intermediate design options.

Peter starts modifying Brad's table on the laptop, and a little while later, he produces Table 1.4.

Table 1.4 Variance of sample mean difference for different designs when treatment 1 is twice as expensive as treatment 2 and the total cost is fixed.

n_1	n_2	$\text{var}(\overline{X}_1 - \overline{X}_2)$	$\sigma_{\overline{X}_1 - \overline{X}_2}$	Efficiency (%)
1	22	1.045	1.022	23.2
2	20	0.550	0.742	44.2
3	18	0.389	0.624	62.4
4	16	0.313	0.559	77.7
5	14	0.271	0.521	89.5
6	12	0.250	0.500	97.1
7	10	0.243	0.493	100.0
8	8	0.250	0.500	97.1
9	6	0.278	0.527	87.4
10	4	0.350	0.592	69.4
11	2	0.591	0.769	41.1

[Peter] Take a look at this.

[Brad] Interesting. Again, the optimal design is not balanced. Its total number of observations is not even an even number.

[Peter, nodding] These results are not quite as dramatic as I would like. The balanced design with eight observations for each treatment is still highly efficient. Yet, this is another example where the balanced design is not the best choice.

[Brad] The question now is whether these examples would be a good start for the book.

[Peter] The good thing about the examples is that they show two key issues. First, the standard design is optimal for at least one scenario, namely, in the scenario where the number of observations one can afford is even, the variances in the two populations are identical and the cost of an observation is the same for both populations. Second, the standard design is often no longer optimal as soon as one of the usual assumptions is no longer valid.

[Brad] Surely, our readers will realize that it is unrealistic to assume that the variances in two different populations are exactly the same.

[Peter] Most likely. But finding the optimal design when the variances are different requires knowledge concerning the magnitude of σ_1^2 and σ_2^2. I don't see where that knowledge might come from. It is clear that choosing the balanced design is a reasonable choice in the absence of prior knowledge about σ_1^2 and σ_2^2, as that balanced design was at least 80% efficient in all of the cases we looked at.

[Brad] I can think of a case where you might reasonably expect different variances. Suppose your study used two machines, and one was old and one was new. There, you would certainly hope the new machine would produce less variable output. Still, an experimenter usually knows more about the cost of every observation than about its variance. Therefore, the example with the different costs for the two populations is

possibly more convincing. If it is clear that observations for treatment 1 are twice as expensive as observations for treatment 2, you have just shown that the experimenter should drop the standard design, and use the unbalanced one instead. So, that sounds like a good example for the opening chapter of our book.

[Peter, laughing] I see you have already lured me into this project.

[Brad] Here is a toast to our new project!

They clink their glasses, and turn their attention toward the menu.

1.3 Summary

Balanced designs for one experimental factor at two levels are optimal if all the runs have the same cost, the observations are independent and the error variance is constant. If the error variances are different for the two treatments, then the balanced design is no longer best. If the two treatments have different costs, then, again, the balanced design is no longer best.

A general principle is that the experimenter should allocate more runs to the treatment combinations where the uncertainty is larger.

2

An optimal screening experiment

2.1 Key concepts

1. Orthogonal designs for two-level factors are also optimal designs. As a result, a computerized-search algorithm for generating optimal designs can generate standard orthogonal designs.
2. When a given factor's effect on a response changes depending on the level of a second factor, we say that there is a two-factor interaction effect. Thus, a two-factor interaction is a combined effect on the response that is different from the sum of the individual effects.
3. Active two-factor interactions that are not included in the model can bias the estimates of the main effects.
4. The alias matrix is a quantitative measure of the bias referred to in the third key concept.
5. Adding any term to a model that was previously estimated without that term removes any bias in the estimates of the factor effects due to that term.
6. The trade-off in adding two-factor interactions to a main-effects model after using an orthogonal main-effect design is that you may introduce correlation in the estimates of the coefficients. This correlation results in an increase in the variances of the effect estimates.

Screening designs are among the most commonly used in industry. The idea of screening is to explore the effects of many experimental factors in one relatively

Optimal Design of Experiments: A Case Study Approach, First Edition. Peter Goos and Bradley Jones.

small study to find the few factors that most affect the response of interest. This methodology is based on the Pareto or sparsity-of-effects principle that states that most real processes are driven by a few important factors.

In this chapter, we generate an optimal design for a screening experiment and analyze the resulting data. As in many screening experiments, we are left with some ambiguity about what model best describes the underlying behavior of the system. This ambiguity will be resolved in Chapter 3. As it also often happens, even though there is some ambiguity about what the best model is, we identify new settings for the process that substantially improve its performance.

2.2 Case: an extraction experiment

2.2.1 Problem and design

Peter and Brad are taking the train to Rixensart, southeast of Brussels, to visit GeneBe, a Belgian biotech firm.

[Brad] What is the purpose of our journey?

[Peter] Our contact, Dr. Zheng, said GeneBe is just beginning to think about using designed experiments as part of their tool set.

[Brad] So, we should probably keep things as standard as possible.

[Peter] I guess you have a point. We need to stay well within their comfort zone. At least for one experiment.

[Brad] Do you have any idea what they plan to study?

[Peter] Dr. Zheng told me that they are trying to optimize the extraction of an antimicrobial substance from some proprietary cultures they have developed in house. He sketched the extraction process on the phone, but reproducing what he told me would be a bit much to ask. Microbiology is not my cup of tea.

[Brad] Likewise. I am sure Dr. Zheng will supply all the details we need during our meeting.

They arrive at GeneBe and Dr. Zheng meets them in the reception area.

[Dr. Zheng] Peter, it is good to see you again. And this must be. . . .

[Peter] Brad Jones, he is a colleague of mine from the States. He is the other principal partner in our firm, Intrepid Stats.

[Dr. Zheng] Brad, welcome to GeneBe. Let's go to a conference room and I will tell you about the study we have in mind.

In the conference room, Brad fires up his laptop, while Dr. Zheng gets coffee for everyone. After a little bit of small talk, the group settles in to discuss the problem at hand.

[Dr. Zheng] Some of our major customers are food producers. They are interested in inhibiting the growth of various microbes that are common in most processed foods. You know, *Escherichia coli*, *Salmonella typhimurium*, etc. In the past they have used chemical additives in food to do this, but there is some concern about the long-term effects of this practice. We have found a strong microbial inhibitor, a certain lipopeptide, in strains of *Bacillus subtilis*. If we can improve the yield of

extraction of this inhibitor from our cultures, we may have a safer alternative than the current chemical agents. The main goal of the experiment we want to perform is to increase the yield of the extraction process.

[Brad] Right.

[Dr. Zheng] The problem is that we know quite a lot already about the lipopeptide, but not yet what affects the extraction of that substance.

[Brad] Can you tell us a bit about the whole process for producing the antimicrobial substance?

[Dr. Zheng] Sure. I will keep it simple though, as it is not difficult to make it sound very complicated. Roughly speaking, we start with a strain of *B. subtilis*, put it in a flask along with some medium, and cultivate it at 37°C for 24 hours while shaking the flask the whole time. The next step is to put the resulting culture in another flask, with some other very specific medium, and cultivate it at a temperature between 30°C and 33°C for some time. The culture that results from these operations is then centrifuged to remove bacterial cells, and then it is ready for the actual extraction.

[Peter] How does that work?

[Dr. Zheng] We start using 100 ml of our culture and add various solvents to it. In the extraction process, we can adjust the time in solution and the pH of the culture.

[Peter] Do you have an idea about the factors you would like to study? The time in solution and the pH seem ideal candidates.

[Dr. Zheng] Yes, we did our homework. We identified six factors that we would like to investigate. We want to look at the presence or absence of four solvents: methanol, ethanol, propanol, and butanol. The two other factors we want to investigate are indeed pH and the time in solution.

[Peter, nodding] Obviously, the response you want to study is the yield. How do you measure it?

[Dr. Zheng] The yield is expressed in milligrams per 100 ml. We determine the yield of a run of the extraction process by means of high-performance liquid chromatography or HPLC.

[Peter] That does not sound very simple either. What is the yield of your current extraction process?

[Dr. Zheng] We have been using methanol at neutral pH for 2 hours and getting about a 25 mg yield per batch. We need something higher than 45 mg to get management interested in taking the next step.

[Brad] That sounds like quite a challenge.

[Peter] How many processing runs can you afford for this study?

[Dr. Zheng] Design of experiments is not an accepted strategy here. This study is just a proof of concept. I doubt that I can persuade management to permit more than 15 runs. Given the time required to prepare the cultures, however, fewer than 15 trials would be better.

[Peter] Twelve is an ideal number for a screening experiment. Using large enough multiples of four allows you to estimate the main effects of your factors independently. You can then save three runs for doing a confirmatory experiment later on. If you think that is a good idea, then I think Brad will have a design for you in less than a minute.

Table 2.1 Brad's design for the extraction experiment at GeneBe.

Run	x_1	x_2	x_3	x_4	x_5	x_6
1	-1	-1	-1	$+1$	-1	-1
2	-1	$+1$	-1	-1	$+1$	-1
3	-1	$+1$	-1	$+1$	$+1$	$+1$
4	$+1$	$+1$	$+1$	-1	-1	-1
5	-1	-1	$+1$	-1	-1	$+1$
6	-1	$+1$	$+1$	$+1$	-1	-1
7	$+1$	$+1$	-1	-1	-1	$+1$
8	$+1$	-1	-1	-1	$+1$	-1
9	$+1$	-1	$+1$	$+1$	$+1$	-1
10	-1	-1	$+1$	-1	$+1$	$+1$
11	$+1$	$+1$	$+1$	$+1$	$+1$	$+1$
12	$+1$	-1	-1	$+1$	-1	$+1$

In a few seconds, Brad turns his laptop so that Dr. Zheng and Peter can see the screen.

[Brad] In generating this 12-run design, I used the generic names, x_1–x_6, for your six experimental factors. I also coded the absence and presence of a solvent using a -1 and a $+1$, respectively. For the factors pH and time, I used a -1 for their low levels and a $+1$ for their high levels.

He shows Dr. Zheng and Peter the design in Table 2.1.

[Dr. Zheng] That was fast! How did you create this table? Did you just pick it from a catalog of designs?

[Brad] In this case, I could have done just that, but I didn't. I created this design ex nihilo by using a computer algorithm that generates custom-built optimal designs.

[Dr. Zheng] That sounds fancy. I was hoping that, for our first project, we could just do something uncontroversial and use a design from a book or a catalog.

[Peter] I have been looking at this design while you two were talking and I would say that, for a two-level 12-run design, this is about as uncontroversial as you can get.

[Dr. Zheng] How so?

[Peter] This design has perfect structure in one sense. Notice that each column only has two values, -1 and $+1$. If we sum each column, we get zero, which means that each column has the same number of -1s and $+1$s. There is even more balance than that. Each pair of columns has four possible pairs of values: $++$, $+-$, $-+$, $--$. Each of these four possibilities appears three times.

[Brad] In the technical jargon, a design that has all these properties is an orthogonal design. In fact, I said earlier that I could have taken this design from a catalog of designs. That is because the optimal design of experiments algorithm in my software generated a design that can be derived from a Plackett–Burman design.

[Peter] A key property of orthogonal designs is that they allow independent estimation of the main effects.