Scott A. Pardo

Empirical Modeling and Data Analysis for Engineers and Applied Scientists

Springer

Empirical Modeling and Data Analysis for Engineers and Applied Scientists

Scott A. Pardo

Empirical Modeling and Data Analysis for Engineers and Applied Scientists

With contributions by Yehudah A. Pardo

Springer

Scott A. Pardo
Ascensia Diabetes Care
Parsippany, NJ, USA

ISBN 978-3-319-32767-9 ISBN 978-3-319-32768-6 (eBook)
DOI 10.1007/978-3-319-32768-6

Library of Congress Control Number: 2016941324

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

Science is about discovery. Discovery is the primary paradigm of science. The primary paradigm of engineering and "applied science" is design. All scientists, whether physicists, biologists, chemists, psychologists, sociologists, anthropologists, economists, geologists, or any other "ists," attempt to discover things. Sometimes, they want to discover the existence of something; sometimes they want to discover how something works; sometimes they want to discover how several things are related; sometimes they want to discover why something exists. Regardless, scientists are in the discovery business. They do not in general want to alter the natural world; they want to understand it. In contrast, the primary paradigm of engineering and applied science is design. Engineers, and those who we will call "applied scientists," want to design things. Clearly, it is important for the engineers and applied scientists, whom we will call EASs, to understand nature and natural phenomena, but understanding is not their goal. Their goal is to exploit nature, hopefully in a beneficial and benevolent manner, in order to make something happen. Thus, the primary goal of the engineer and applied scientist is design.

Statistics, as a discipline, is mostly oriented toward the discovery paradigm. Statistics courses emphasize creating predictive models or classificatory models, either predicting nature or classifying individuals. Most commonly, we hope to reject the hypothesis of no effect, in favor of discovering an effect. It seems that often statistics is used to prove or disprove the existence of some phenomenon, as opposed to aiding in the design of a product or process. This is not to say that statistical methods cannot be used, or are never used, to help design something. Chemical engineers may use designed experiments to optimize a process; manufacturing engineers may use experimental data to optimize the operation of a machine; industrial engineers might use data to determine the optimal number of operators required in a manual assembly process. This text is about gathering and analyzing empirical observations (data) in order to aid in making design decisions. The EAS may believe that experimentation is unnecessary for designing. He or she might believe that design decisions should be made without any empirical observation and that experimentation is only useful for verifying or validating designs. Every electrical engineer knows that $V = IR$, but what happens to V if both I and R have some random components? What about the ideal gas law, $P = k\frac{T}{V}$? There seems to be no need for empirical data when applying these laws. The formulas and equations learned in an elementary physics course may take on new meaning when accounting for probabilistic variation. Also, there are many design situations where no simple equation exists. This text is meant to speak to the EAS, and hopefully motivate her or him to experiment, with the design objective in mind.

Much of the discussion in this book is about models. Models are by definition incorrect. The question is not whether the model truly represents reality, but rather whether the model adequately

represents reality with respect to the problem at hand. Many of the ideas presented will focus on how to gather data in the most efficient way possible in order to construct an adequate model.

The statistical methods presented are not new. In general, the techniques and concepts introduced in this book are meant to stimulate the reader's imagination and not meant to be the definitive answers to problems. Certainly, the ideas presented are not an exhaustive list. The authors hope that this book will present a variety of design situations familiar to many engineers and applied scientists and inspire the reader to incorporate experimentation and empirical investigation into the design process.

Software is integrally linked to statistical analyses. Examples in this book have been worked using several packages/languages/programs, notably SAS, R, JMP, Minitab, and MS Excel. It is the authors' belief that there is no "best" software in general. All packages and languages have advantages and disadvantages. The point of using several types of software was simply to demonstrate that no one package or language is best overall. This text is not a primer on software, however. It is assumed that the reader has familiarity with some data analysis software.

This material can be used at the advanced undergraduate or first-year graduate level. The students who would most benefit from this book are those studying engineering or applied science. The student would benefit greatly from some accompanying laboratory work. While fully worked examples are given in every chapter, there is no teacher like hands-on experience. Most of the chapters in this book are subjects that are covered in an entire book by itself. The goal is to introduce the student to ideas about empirical investigation in such a way as to motivate him or her to use experimentation as an aid to design.

The authors encourage instructors to assign the students practical experience in conducting experiments, making measurements and observations, and analyzing their data. Ideally, the student should use data that are intrinsically meaningful to him or her, such as experimental data associated with a thesis or dissertation. The fundamental learning objective of this book is for the reader to understand how experimental data can be used to make design decisions and to be familiar with the most common types of experimental designs and analysis methods.

Although the text includes introductory chapters in probability and statistics, it would greatly help the student to have already been exposed to those subjects, as well as some linear algebra.

We must make a small apology about the letter "*p*." We use this letter to symbolize probability, numbers of parameters in a model, and powers of ½. It can be a little confusing. At least the reader is warned.

A brief word about data-intensive modeling methods, such as artificial neural networks and fuzzy algorithms, is appropriate. This is brief, because those methods are not mentioned at all in the text. While valuable and important, they could have and have had entire texts devoted to those techniques. This text will focus on methods that can be used with "small" data sets, generally gathered in a designed experiment.

How to Use This Book as a Text

This book could be used as a text for a course titled something along the lines of "Statistical Methods for Engineers and Applied Scientists," "Experimentation in the Design Process," or "Using Empirical Data to Aid in the Design of Products and Processes." It could also provide students some more in-depth discussion of statistical methods discussed in a Design for Six Sigma course. The first seven chapters are largely about factorial experimentation, although the material in Chap. 3 on measurement systems does not traditionally appear in experimental design texts. The remaining chapters might be called "special topics in data analysis," and much of that material involves application of experimental designs. The book is intended to stimulate students to engage in empirical investigation

as part of their design process. It is not a text about engineering design, nor is it strictly an experimental design text. There are many topics in experimental design and analysis that are not included (e.g., split-plot designs, One-way ANOVA, partially balanced incomplete blocks, and the method of steepest ascent), and virtually no discussion about engineering design, per se. Rather, it is intended to help the student understand how empirical investigation and empirical models could be used to aid in design. If students had previously taken a course in the elements of probability and statistical theory, the first two chapters could be skipped. Otherwise, the authors suggest covering Chaps. 1 and 2 in the first week and one chapter each week thereafter. Some of the chapters, notably Chaps. 11 (Reliability) and 15 (Robust Design), might require more time than 1 week. Of course, the instructor should use her or his discretion in including additional materials, excluding some of the text, or the timing of coverage for any of the text's material.

Parsippany, NJ, USA Scott A. Pardo

Acknowledgments

S.P. would like to acknowledge the contributions of his wife and partner, and Y.P. his mother, to the completion of this book. She provided insights and suggestions that greatly enhanced the structure and content of the text, making it more complete and useful than it would have been without her help. They also acknowledge the contributions of Michael A. Pardo and Jeremy D. Pardo, sons, brothers, and collaborators, whose descriptions of their own scientific endeavors, and questions about our intentions and thoughts, redirected our thinking and helped us broaden our vision of the "applied scientist." Finally, we would like to thank Dr. Rezi Zawadzki for her reviews of the manuscript, comments, and encouragement.

Contents

About the Authors

Scott A. Pardo has been a professional statistician since 1980. He has worked in a wide variety of industrial contexts, including the US Army Information Systems Command, satellite systems engineering, pharmaceutical development, and medical devices. He is a Six Sigma Master Black Belt, is an Accredited Professional Statistician (PStat™), and holds a Ph.D. in Industrial and Systems Engineering from the University of Southern California.

Yehudah A. Pardo is a Ph.D. student in biomedical engineering at Cornell University. His research interests involve cell-free protein expression and novel biomaterial platforms for modeling tissues and disease. Previously, his research work focused on extracting the photosynthetic machinery from cyanobacteria for the development of a biological photovoltaic cell. He holds a B.S. in Bioengineering from Binghamton University, State University of New York.

Chapter 1
Some Probability Concepts

Probability begins with the ideas of "sample space" and "experiment". An experiment is the observation of some phenomenon whose result cannot be perfectly predicted a priori. A sample space is the collection of all possible results (called outcomes) from an experiment. Thus, an experiment can be thought of as the observation of a result taken from a sample space. These circular-sounding definitions may be a little annoying and somewhat baffling, but they are easily illustrated. If the experiment is to observe which face of a six-sided die lands up after throwing it across a gaming table, then the sample space consists of six elements, namely the array of 1, 2, 3, 4, 5, or 6 dots, as they are typically arrayed on the faces of a six-sided die. An event is a set of outcomes. So, for example, the set $A = \{1, 3, 5\}$ could represent the event that an odd number of dots shows up after throwing a six-sided die. Events have probabilities associated with them. For discrete events, such as in the die-throwing experiment, the probability is the number of outcomes contained in the event set divided by the total number of outcomes possible. So, the probability of event A as previously defined is:

$$\Pr\{A\} = \frac{\#outcomes\ in\ A}{\#outcomes\ possible} = \frac{3}{6}$$

Sample spaces need not be so discrete or finite; they can be continuous and infinite, in that they can have an infinite number of outcomes. For example, if a sample space consists of all possible initial voltages generated by LiI batteries made in a battery manufacturing plant, then it would have an infinite (but bounded) number of possible outcomes.

A random variable is a mapping from a sample space into (usually) some subset of the real numbers (possible over the entire real line). Think of the random variable as a "measurement" taken after the experiment is performed. Thus, the number of dots in the array showing after the die is cast, or the voltage as measured by a volt meter, would be random variables. There are two basic classes of random variables, discrete and continuous. Discrete random variables are mapped from the sample space to (possibly infinite) subsets of integers, and continuous random variables are mapped to (possibly infinite) subsets of real numbers. The die example is discrete, and the voltage example is continuous.

Every random variable has a probability distribution function that describes the chances of observing particular ranges of values for the random variable. In the case of discrete random variables, it also makes sense to talk about the probability of an experiment resulting in a particular value, e.g., the probability that the number of dots in the die array showing is 4. For continuous variables, it makes sense to talk about the probability of obtaining a value in a "small" range, but the probability of obtaining a particular value is 0. This is not to say that particular values of continuous

S.A. Pardo, *Empirical Modeling and Data Analysis for Engineers and Applied Scientists*,
DOI 10.1007/978-3-319-32768-6_1

random variables are never observed or measured; it just means that we do not have the ability to predict a particular value with any non-zero measure of uncertainty.

A probability cumulative distribution function (cdf) describes the probability that a random variable is less than or equal to a particular value. We will use capital letters to represent the random variable, and lower case letters to represent particular values. If X is a random variable, then the cdf for X is symbolized as:

$$F_X(x) = Pr\{X \leq x\}$$

In the case of discrete random variables, this function is a sum of probabilities for particular values, $p(x_k)$, up to and including the value x:

$$F_X(x) = \sum_{x_k \leq x} p(x_k)$$

The function $p(x_k)$ is referred to as the probability mass function (pmf). In the case of continuous random variables, the summation is replaced with and integral, and the discrete probability mass function is replaced with something called a probability density function, or pdf (usually; there are some more or less degenerate cases where a density function does not exist), $f(x)$, which defines the probability that the random variable would have values observed in a small interval, dx:

$$f_X(x)dx = Pr\{x - dx \leq X \leq x + dx\}$$

So the cdf is:

$$F(x) = \int_{-\infty}^{x} f_X(\xi)d\xi$$

In general, the probability mass functions and density functions are defined in terms of parameters that give these functions their particular characteristics. This book involves several special classes of density functions and their associated parameters.

There are some special characteristics of random variables called moments. We will only be concerned with two such characteristics, called expectation (or mean) and variance (and its square root, called standard deviation). The expectation of a random variable is given by:

$$E\{X\} = \mu = \begin{cases} \sum_k x_k p(x_k) \\ \int_{-\infty}^{+\infty} \xi f(\xi)d\xi \end{cases}$$

The sum is for discrete random variables, and the integral for continuous. The expectation is like the center of gravity for the random variable, if one thinks about the density function describing the distribution of mass over a beam. The variance is:

$$E\left[(X-\mu)^2\right] = \sigma^2 = \begin{cases} \sum_k (x_k - \mu)^2 p(x_k) \\ \int_{-\infty}^{+\infty} (\xi - \mu)^2 f(\xi)d\xi \end{cases}$$

Table 1.1 Some probability density and mass functions

Name	Parameters	Density or mass function	Range of values
Normal	μ, σ	$\frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$	$-\infty < x < +\infty$
Gamma	n, λ	$\frac{\lambda^n}{\Gamma(n)} x^{n-1} \exp(-\lambda x)$	$x > 0$
Chi-squared	ν	$\frac{(1/2)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2}-1} \exp\left(-\frac{1}{2}x\right)$	$x > 0$
Student's t	ν	$\frac{\Gamma\left(\frac{1}{2}(\nu+1)\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{1}{2}\nu\right)} \left[1 + \frac{x^2}{\nu}\right]^{-\frac{(\nu+1)}{2}}$	$-\infty < x < +\infty$
F	ν_1, ν_2	$\frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{1}{2}\nu_1\right)\Gamma\left(\frac{1}{2}\nu_2\right)} \frac{x^{\left(\frac{\nu}{2}\right)-1}}{(1+x)^{(\nu_1+\nu_2)/2}}$	$x > 0$
Poisson	λ	$\frac{\lambda^k e^{-\lambda}}{k!}$	$k = 0, 1, 2, \ldots$
Binomial	n, p	$\binom{n}{k} p^k (1-p)^{n-k}$	$k = 0, 1, 2, 3, \ldots n$
Beta	α, β	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$	$0 < p \leq 1$

Again, the summation and integral are for discrete and continuous random variables, respectively. The square root of σ^2 is called the standard deviation, and it is useful in making probability calculations with the normal probability distribution. Note that the expected value of any function of the form $(X - \mu)^r$ is called the rth moment about the mean. That is, the rth moment about the mean is:

$$E[(X-\mu)^r] = \begin{cases} \sum_k (x_k - \mu)^r p(x_k) \\ \int_{-\infty}^{+\infty} (\xi - \mu)^r f(\xi) d\xi \end{cases}$$

It turns out that most of the time, parameters of distributions can be expressed in terms of these moments.

Table 1.1 shows the parametric forms of density and mass functions for several special random variables referred to in the text.

Some important concepts used throughout the text are mutual exclusivity, independence, conditional probability and conditional expectation. If A and B are two events, the joint probability is the probability that after the experiment is performed, both A and B would have occurred. Once again using the die-throwing experiment, suppose $A = \{$odd number of dots showing$\}$ $B = \{$number of dots showing is less than 4$\}$. The joint probability would be the number of outcomes in the intersection of the two sets A and B divided by the total number of outcomes, namely:

$$Pr\{A \cap B\} = \frac{\#\{1,3,5\} \cap \{1,2,3\}}{\#\{1,2,3,4,5,6\}} = \frac{2}{6}$$

The probability that either event A occurs, event B occurs, or both event A and B occur is the probability of the union of the two sets A and B. In general:

$$\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A \cap B\}$$

Two events are called mutually exclusive if:

$$\Pr\{A \cap B\} = 0$$

Two events are called independent if:

$$\Pr\{A \cap B\} = \Pr\{A\}\Pr\{B\}$$

Conditional probability is the probability of an event given that another event is known to have occurred.

The idea is that the probability of an event, given that some particular condition is known to exist, depends on the particular condition. As an example, the probability that the air temperature at the earth's surface, given that it is measured inside the arctic circle in January, is between 35 and 45 °C, is not the same as the probability of obtaining a temperature between 35 and 45 °C, given temperature is measured in Death Valley, CA, in the same time period. The notation we will use for conditional probability is:

$$Pr\{A|B\}$$

which is read, "the probability of event A given that B is known to have occurred", or more simply, "the probability of A given B".

The conditional probability is calculated as:

$$\Pr\{A|B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}}$$

Manipulating the formula for conditional probability gives an expression for the joint probability in terms of conditional probabilities:

$$Pr\{A \cap B\} = Pr\{A|B\}Pr\{B\} = Pr\{B|A\}Pr\{A\}$$

There is a generalization of this relationship, called Bayes' Theorem. Suppose there are a set of events, $B_1, B_2, \ldots, B_k$ that are mutually exclusive and they partition the sample space, S, i.e.:

$\{B_i\} \cap \{B_j\} = \phi = \{\}$ = the empty set, $i \neq j$, and $\cup_{i=1}^{k}\{B_i\} = S$

Then if A is some other event, then

$$\Pr\{A\} = \sum_{i=1}^{k}\Pr\{A \cap B_i\} = \sum_{i=1}^{k}\Pr\{A|B_i\}\Pr\{B_i\}$$

Thus, Bayes' theorem states that

$$\Pr\{B_j|A\} = \frac{\Pr\{A|B_j\}\Pr\{B_j\}}{\sum_{i=1}^{k}\Pr\{A|B_i\}\Pr\{B_i\}}$$

This theorem, and its continuously-valued analog, is particularly useful when A represents observations (data), the B_i represent different possible values (or sets of values) for some parameter, $\Pr\{A|B_i\}$ is the likelihood of observing A if the parameter is equal to (or in the set) B_i, and $\Pr\{B_i\}$ is the degree to which it is believed, prior to getting data A, that the parameter equals (or is in the set) B_i.

Expected values can also be conditional. For example, the average duration of daylight, given that the location is New York City in June, will not be the same as the average duration of daylight, given the location is Melbourne, Australia in June. The notation is:

$$E[Y|x]$$

where Y is a random variable, and x is a known condition, which in turn could be a particular value of another random variable, or the value of some parameter. Formally, conditional expectation is defined in terms of conditional probability mass or density functions. That is, perhaps the density of variable Y depends of the value of another variable, X. If $f(y|X = x)$ is the conditional density of Y given $X = x$, then the conditional expectation of Y given $X = x$ is:

$$E[Y|X = x] = \int_{-\infty}^{+\infty} yf(y|X = x)dy$$

Two random variables, X and Y, are said to have a joint probability distribution, with a joint cumulative distribution function $F(x,y) = Pr\{X \leq x \text{ AND } Y \leq y\}$. If it exists, the two variables have a joint density function:

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

The two variables are said to be independent if:

$$F(x, y) = F_X(x)F_Y(y)$$

$F_X(x)$ and $F_Y(y)$ are the respective cumulative distribution functions of each random variable. These one-variable cdfs are referred to as marginal cdfs. Similarly, if the joint density function exists, then when X and Y are independent:

$f(x,y) = f_X(x)\, f_Y(y)$. The conditional density of Y given $X = x$ is expressed in terms of the joint and marginal densities:

$$f_{Y|_x}(y|x) = \frac{f(x, y)}{f_X(x)}$$

Key Points

- Probability is a mapping from sets, called events, into the interval [0,1].
- Random Variables (RVs) are mappings from sample spaces into (usually) the real numbers.
- RVs can either be discrete or continuous.
- RVs have cumulative distribution functions (cdfs); discrete RVs have probability mass functions; continuous RVs usually have probability density functions.
- CDFs often have parametric forms.
- Bayes' Theorem provides a convenient way of using data to update the uncertainty about a distribution's parameters.

Exercises and Questions

1. A production lot of 100 million plastic beads consists of 60 % red, 20 % white, and 20 % blue beads. Five percent of the red beads are defective and 2.5 % of the white beads are defective. The overall percent defective beads is 7 %. What percent of the blue beads are defective? A: 17.5 % (Hint: express Pr{Defective} in terms of the conditional probabilities of defective given bead color)
2. The expected value of the weight of a seed from a hybrid corn plant is 3.0 g, and the variance is 0.01. Assuming that seed weight is normally distributed, what is the probability that a seed will weigh between 2.75 and 3.25 g? A: ~0.98758
3. Are there any situations where the probability of an event is not conditional?
4. How do you interpret the statement: “There is a 70 % chance of rain tomorrow.”?

Chapter 2
Some Statistical Concepts

This book is concerned with making inferences about parameters of probability distribution functions. An inference is a generalization made from some specific observations. The specific observations are the data; the generalization is about the values of the parameters. The data are presumed to be a (relatively) small subset of values obtained, measured, or observed in some way from a larger population (sample space). Generally, the parameters are unknown. What we have instead are sample statistics, which are functions of the data. These statistics are themselves random variables, in that every new subset of values from the population yields potentially at least a new value for the statistic. As a result, the sample statistic also has a sample space associated with it, and a probability distribution function as well. The probability distribution function for a sample statistic is often referred to as a sampling distribution function (Meyer 1970). The form of the sampling distribution usually depends on the formula for the statistic, and the distribution function of the random variable for which the data constitute a subset of values or observations.

One common situation is to make inferences about the expected value of a random variable having a normal probability density, i.e.: $E(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} x e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \mu$ (it is a convenient coincidence that the expected value of a normally distributed random variable happens to be one of its parameters)

The problem is that both μ and σ, the two parameters for the normal distribution, may not have known values. We can only infer something about this expected value based on a finite subset of values from this normally distributed population. Let $x_1, x_2, x_3, \ldots, x_n$ represent the values of this finite subset, called a sample. We can compute two sample statistics:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \text{ and } s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

These represent sample estimates for the population parameters μ and σ. These estimates are referred to as "point" estimates, in that they are single values, and not a range or interval.

An inference would be made if we wanted to know (infer) that the expected value of this particular normal distribution was equal to a particular value or not.

There is some formalism, called hypothesis testing, concerning inference. The notion is that we do not know the value of a parameter, but perhaps we would like to know specifically if the parameter either equals a particular value or if it falls in some particular range. Hypotheses always come in pairs; the "null" hypothesis, usually symbolized as H_0 (hence the term "null") and the alternative, H_1, which is in some sense the logical negation of the null. The idea is that a test statistic, such as $\overline{x}$, would, if the null hypothesis is true, have a value that would fall within some interval with some a

S.A. Pardo, *Empirical Modeling and Data Analysis for Engineers and Applied Scientists*,
DOI 10.1007/978-3-319-32768-6_2

priori determined probability, which we will call $1 - \alpha$. The probability, α, is the probability of obtaining a value of the statistic outside the pre-specified range even though the null hypothesis is true; this is called Type I risk. Since we do not know if the null is true or not, we only decide to believe it (accept) or disbelieve it (reject). Often, the alternate hypothesis can be true under an infinite number of possible alternative values for the parameter of interest. For example, if we were interested in testing the hypothesis that a population mean, μ, was either equal to a specific value, say μ_0, or that μ is not equal to μ_0, the pair of hypotheses might look like:

H_0: $\mu = \mu_0$ vs. H_1: $\mu \neq \mu_0$.

Of course, there are an infinite number of possible values for μ that would make the null hypothesis false. The probability of rejecting the null hypothesis when it is false is called power. The power, usually symbolized as $1 - \beta$, is a function of alternate values of μ (i.e., values other than μ_0). The value of β is referred to as Type II risk. Usually, a curve of power as a function of alternate values of the parameter is constructed. Oddly enough, this curve is referred to as a "power curve".

If we "hypothesize" about whether the expected value is equal to μ_0, i.e., some specific value, we form yet another sample statistic:

$$t = \frac{\sqrt{n}(\overline{x} - \mu_0)}{s}$$

This statistic, if the expected value of the random variable X actually is equal to μ_0, has a sampling distribution called Student's t with a parameter called degrees of freedom (df) equaling the convenient (and known) value $n - 1$. The inference to be made is whether it is believable that the expected value of X is equal to μ_0 or not. If the value of the statistic t falls within a "reasonable" range we would expect (say a range that covers 95 % of values for a random variable having a Student's t distribution with $df = n - 1$). In other words, if $p = \frac{\alpha}{2}$, $\alpha = 0.05$, and t_p represents the $100p$ percentile of this Student's t distribution, we would expect the sample statistic to fall somewhere between $t_p = t_{0.025}$ and $t_{1-p} = t_{0.975}$ with probability 0.95 (95 %). So the inferential rule for this statistic could be the following:

If the sample statistic, t, falls in the interval ($t_{0.025}$,$t_{0.975}$) then we are willing to believe that the expected value of the random variable we were sampling is equal to μ_0. Otherwise, we will not believe it. Alternatively, we can calculate the probability that, given the null hypothesis is true, a Student's t variable would be greater than (in absolute value) the computed test statistic, t. This probability is called a p-value. If the p-value is lower than some pre-specified level, say 0.05, then we would reject the null hypothesis H_0 in favor of the alternate, H_1. The power is computed by calculating the probability that the t statistic would fall outside the range ($t_{0.025}$,$t_{0.975}$) when in fact μ equals to some values other than μ_0. If μ actually equals μ_0, then the power is equal to 0.05 (i.e., α). In the case of the t statistics, the power is calculated using probabilities from something called a non-central t distribution (Bickel and Doksum 2007). The non-central t (which we will symbolize as T') distribution has an extra parameter, called the non-centrality parameter (ncp), which would equal in this case:

$$ncp = \frac{\sqrt{n}(\mu_a - \mu_0)}{\sigma}$$

The value of μ_a is varied, and the power is the probability calculated using the non-central t:

$$Pr\left\{T' < t_{0.025}\ OR\ T' > t_{0.975} \middle| ncp = \frac{\sqrt{n}(\mu_a - \mu_0)}{\sigma}\right\}$$

The values $t_{0.025}$ and $t_{0.975}$ are percentiles of the usual (central) t distribution. As sample size increases, the ability to "detect" departures from H_0 (i.e., reject the null hypothesis) increases. There are virtually an infinite number of possible test statistics. However, there are only a few for which the distribution of the statistic, given some null hypothesis, is relatively easy to use to make probability calculations. In this text, we will mostly encounter t statistics and F statistics.

A concept closely related to hypothesis testing is the confidence interval. Unlike point estimates, confidence intervals are ranges of values that have some stated probability (confidence level) of containing the actual value of parameters. Confidence intervals are computed using data, and they can be constructed for any parameter. The trick is that the confidence intervals of some parameters are more difficult to construct than others. Nowadays, with modern computing capabilities, it is relatively easy to construct confidence intervals for parameters that can be expressed in terms of expectations. The method of boot-strapping (Efron 1982) is particularly useful in computing confidence intervals. In some important cases, however, confidence intervals have relatively simple closed-form expressions. In particular, the $100(1 - 2\alpha)$ % confidence interval for the expected value, μ, given the sample mean, $\overline{x}$, and sample standard deviation, s, is:

$$\overline{x} \pm t_{1-\alpha}(n-1)\frac{s}{\sqrt{n}}$$

The constant $t_{1\text{-}\alpha}(n - 1)$ is the $100(1 - \alpha)$ percentile of s Student's t distribution with $n - 1$ degrees of freedom. The value:

$$\frac{s}{\sqrt{n}}$$

is the sample estimate of the standard error of the mean, which is the standard deviation of sample means. Note that as the sample size, n, increases, the width of the confidence interval decreases. Thus increasing sample size increases the precision of estimation.

One of the nice things about the confidence interval formula for the mean is that it actually yields a truly 95 % confidence interval for μ, regardless of how the random variable X is distributed, at least as the sample size gets "big enough". This fact is due to something called the central limit theorem. Of course, "big enough" depends on how close you need to be to having $100(1 - 2\alpha)$ % confidence. Even smaller sample sizes (say 10 or 20) are probably adequate for insuring that the confidence level is really about $100(1 - 2\alpha)$ %. In general, and the sample size increases, the width, or imprecision, of the confidence interval decreases. It is important to understand that increased sample size does not increase the confidence level; rather, increasing sample size decreases the width of the interval. Two intervals, having the same confidence level, could have very different widths. Suppose that with a sample size of $n = 5$, the experimenter obtained a sample mean of $\overline{x} = 10$ and a sample standard deviation of $s = 2$. With this sample size, and $\alpha = 0.025$, $t_{1-\alpha} = t_{0.975} \approx 2.78$. Then the $100(1 - 2(0.025))\ \% = 95$ % confidence interval for the mean would be:

$$\overline{x} \pm t_{0.975}\frac{s}{\sqrt{n}} \approx 10.0 \pm 2.78\frac{2}{\sqrt{5}} \approx 10.0 \pm 2.49 = (7.51, 12.49)$$

Now suppose the experimenter obtained $\overline{x} = 10$ and $s = 2$, but $n = 35$. Then $t_{0.975} \approx 2.03$ and the 95 % confidence interval would be approximately:

$$\overline{x} \pm t_{0.975}\frac{s}{\sqrt{n}} \approx 10.0 \pm 2.03\frac{2}{\sqrt{35}} \approx 10.0 \pm 0.69 = (9.31, 10.69)$$

Both of these are 95 % confidence intervals, and both are based on identical values for the sample statistics. However, the interval with $n = 35$ is considerably narrower in width that the interval with $n = 5$.

The confidence level is statement about the interval, and not about the parameter. That is, the parameter is assumed to have some unknown value. The EAS constructs an interval that has a 100 $(1 - 2\alpha)$ % chance of "capturing" the true value of the parameter. The precision of that interval depends mostly on the sample size.

For an excellent coverage of probability and statistical topics, see Meyer (1970).

A Brief Note on Sample Size Estimation

The topic of how to determine an appropriate sample size is in fact worthy of an entire book. Generally, sample size is chosen to either provide a particular probability to reject a null hypothesis when the truth departs from the null assumption by some specific quantity, or to provide a confidence interval of some pre-specified width. For a much more complete discussion of how to choose a sample size, see Desu and Raghavarao (1990).

Key Points

- Statistics are computations made using empirical observations, and are used to estimate parameters of a population. Any computation with data is referred to as an "estimate".
- Statistics are themselves random variables, and as such have distributions, called "sampling distributions".
- Sampling Distributions are used to make inferences about population parameters, through hypothesis tests and confidence intervals.
- A confidence interval can be constructed for any parameter; some confidence intervals are easier to compute than others.

Exercises and Questions

1. The 95 % confidence interval, based on a sample of n parts, for mean length of a steel rivet is (0.245 cm, 0.255 cm). Does this mean that there is a 95 % probability that the true mean length is between 0.245 and 0.255 cm?
2. How is a confidence interval affected by increasing the sample size? Consider the confidence interval formula for the expected value.
3. What does the width (difference between upper and lower limits) of a confidence interval tell you?
4. Will a sample standard deviation shrink to zero as the sample size increases?

Chapter 3
Measurement Systems Analysis

The EAS bases most of his or her decisions about design parameters on measurements, so it is imperative that those measurements be trustworthy. There are two categories of measurement quality we will discuss, namely accuracy and precision. Accuracy is the degree to which the measurement differs from the truth, on the average. Precision is the degree to which the measurements vary from instance to instance of measuring the same unit or item. These definitions would probably be better termed "inaccuracy" and "imprecision", but we will use the terms "accuracy" and "precision" to more or less mean inaccuracy and imprecision. In order to assess accuracy, the true value of the dimension or performance parameter for each part (or item being measured) in the sample must be known. In lieu of knowing the true value, a more trustworthy "reference method" might be used to provide a more accurate and precise measurement against which the measurement system to be evaluated will be compared. This chapter will treat both the case where only precision can be evaluated (no reference) and where accuracy may also be evaluated (in comparison to a reference method result).

No Reference Results Available

Initially we will focus on precision. The first question to ask is "how precise is precise enough?" Consider the problem of measuring items to determine whether or not the dimension or parameter of interest is within specified limits. The simplest case is when the measurement is continuously-valued, and the dimension or parameter has both a lower and upper specification limit, and when the measurement is non-destructive, so that items (which we will refer to as "parts") may be measured multiple times without being altered in any meaningful way. We are hoping that the amount of variability when an operator measures a particular part multiple times is small compared to the variability between parts. Furthermore, we are also hoping that regardless of which operator uses the measurement device (which we will call "gauge"), the differences between operators is small compared to the differences or variation between parts. Suppose that L and U represent the lower and upper specification limits, respectively, for the dimension or parameter of interest. That means if any part's dimension is anywhere in the interval $[L, U]$, then that part would be considered acceptable for use. Suppose further that the variation in measurements made on a single part by a single operator is normally distributed, with standard deviation σ_w. Assuming further that on the average, the measurements are at the midpoint of the specification range, $\frac{U+L}{2}$, then the range $6\sigma_w$ would contain approximately 99.73 % of the measurements made on a single part by a single operator. Since the

S.A. Pardo, *Empirical Modeling and Data Analysis for Engineers and Applied Scientists*,
DOI 10.1007/978-3-319-32768-6_3

Fig. 3.1 Combinations of decisions and states

	Actual State	
Measurement Decision	Part is In-Spec	Part is Out-of-Spec
Part is In-Spec (-)	P_I	Pr{False "-"}
Part is Out-of-Spec (+)	Pr{False "+"}	P_O

range $U - L$ is the range of acceptable values for the dimension, then it would be desirable for the ratio:

$$\frac{6\sigma_w}{U-L}$$

to be small. Conversely, we would want the reciprocal, which is the "capability index" called C_p, to be large. Keep in mind that in this case, σ_w represents the variability in the measurement of a single part, and not the variability between parts. If C_p were equal to 1, then for a part whose dimension is exactly at the midpoint of the specification range, there would be approximately a 99.73 % chance that the measurement would be within the limits. The smaller the within-part, within operator variation, the larger the value of C_p would get. Values of C_p greater than 1 are desirable.

If a part's dimension is equal to either U or L, then there would be a fairly high chance that the part's measurement would be outside the limits. The problem is to assess the conditional probabilities:

$P_I = Pr\{\text{measurement is within (U,L)} \mid \text{a part's dimension is truly within (U,L)}\}$
$P_O = Pr\{\text{measurement is outside (U,L)} \mid \text{a part's dimension is truly outside (U,L)}\}$

Symbolize detection of an out-of-spec part with "+", and no detection with a "−". Figure 3.1 illustrates the four possible combinations of measurement-based decisions and actual true states.

Clearly we desire both P_I and P_O to be "high" probabilities (somewhere in the range of 95–99.99 %). Suppose X represents the measurement, and Y represents the true value of the part's dimension. If we assume that both have normal distributions, and that the expected value of X given $Y = y$ is the true value of the dimension, and the average part's dimension is truly μ_B, then we can represent the probability $Pr\{L \le X \le U | Y = y\}$ mathematically:

$$Pr\{L \le X \le U | Y = y\} = \int_L^U \frac{1}{\sqrt{2\pi}\sigma_w} e^{-\frac{1}{2}\left(\frac{x-y}{\sigma_w}\right)^2} dx$$

The joint probability $Pr\{L \le X \le U, L \le Y \le U\}$ is given by:

$$Pr\{L \le X \le U,\ L \le Y \le U\} = \int_L^U \frac{1}{\sqrt{2\pi}\sigma_B} e^{-\frac{1}{2}\left(\frac{y-\mu_B}{\sigma_B}\right)^2} \int_L^U \frac{1}{\sqrt{2\pi}\sigma_w} e^{-\frac{1}{2}\left(\frac{x-y}{\sigma_w}\right)^2} dxdy$$

The probability, P_I, is then:

$$Pr\{L \le X \le U | L \le Y \le U\} = \frac{\int_L^U \frac{1}{\sqrt{2\pi}\sigma_B} e^{-\frac{1}{2}\left(\frac{y-\mu_B}{\sigma_B}\right)^2} \int_L^U \frac{1}{\sqrt{2\pi}\sigma_w} e^{-\frac{1}{2}\left(\frac{x-y}{\sigma_w}\right)^2} dxdy}{\int_L^U \frac{1}{\sqrt{2\pi}\sigma_B} e^{-\frac{1}{2}\left(\frac{y-\mu_B}{\sigma_B}\right)^2} dy}$$

Assuming that the events $\{Y < L\}$ and $\{Y > U\}$ are mutually exclusive, then we can partition the probability P_O into two additive terms:

$$Pr\{X < L \; OR \; X > U | Y < L \; OR \; Y > U\} = Pr\{X < L \; OR \; X > U | Y < L\} Pr\{Y < L\} + Pr\{X < L \; OR \; X > U | Y > U\} Pr\{Y > U\}$$

The left-hand side expression is a slight abuse of notation, but we want to emphasize the dependence upon conditions involving Y. We can make some further simplifying assumptions that if $Y < L$, it is virtually impossible for $X > U$, and similarly it is of negligible probability that if $Y > U, X < L$. The approximation is then:

$$P_O = Pr\{X < L \mid Y < L\} Pr\{Y < L\} + Pr\{X > U \mid Y > U\} Pr\{Y > U\}$$

The two terms are given by:

$$Pr\{X < L | Y < L\} \Pr\{Y < L\} = \int_{-\infty}^{L} \frac{1}{\sqrt{2\pi}\sigma_B} e^{-\frac{1}{2}\left(\frac{y-\mu_B}{\sigma_B}\right)^2} \int_{-\infty}^{L} \frac{1}{\sqrt{2\pi}\sigma_w} e^{-\frac{1}{2}\left(\frac{x-y}{\sigma_w}\right)^2} dxdy$$

and

$$Pr\{X > U | Y > U\} \Pr\{Y > U\} = \int_{U}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_B} e^{-\frac{1}{2}\left(\frac{y-\mu_B}{\sigma_B}\right)^2} \int_{U}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_w} e^{-\frac{1}{2}\left(\frac{x-y}{\sigma_w}\right)^2} dxdy$$

Even in their most simple forms, these probabilities have no closed form, and are fairly complex for computation. As a heuristic, rule-of-thumb, method for determining how much variability and error to tolerate in a measurement system, first consider the two conditional probabilities:

$$Pr\{X < L | Y = L\} = \int_{-\infty}^{L} \frac{1}{\sqrt{2\pi}\sigma_w} e^{-\frac{1}{2}\left(\frac{x-L}{\sigma_w}\right)^2} dx$$

$$Pr\{X > U | Y = U\} = \int_{U}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_w} e^{-\frac{1}{2}\left(\frac{x-U}{\sigma_w}\right)^2} dx$$

We are assuming that the conditional expectation of X given $Y = y$ is y. Both of these probabilities are in fact equal to 50 %. This is not helpful. That is, the sum of the two probabilities is 1.0:

$$Pr\{X < L | Y = L\} + Pr\{X > U | Y = U\} = 0.50 + 0.50 = 1.0$$

This fact is fairly obvious, and follows from the symmetric nature of the normal density function.

What we need is another parameter, call it ϵ, such that:

$$Pr\{X < L | Y = L + \epsilon\} = \int_{-\infty}^{L} \frac{1}{\sqrt{2\pi}\sigma_w} e^{-\frac{1}{2}\left(\frac{x-L-\epsilon}{\sigma_w}\right)^2} dx = p$$

and:

$$Pr\{X > U | Y = U - \epsilon\} = \int_{U}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_w} e^{-\frac{1}{2}\left(\frac{x-U+\epsilon}{\sigma_w}\right)^2} dx = p$$

where p is a sufficiently small probability. In other words, if the true dimension is inside the specification range by some small, pre-specified amount ϵ, we want only a small chance that the measurement would fall outside the range (L, U). The probabilities lead to the following equations:

$$\frac{-\epsilon}{\sigma_w} = z_p$$

and

$$\frac{\epsilon}{\sigma_w} = z_{1-p}$$

where z_p is the z-score for probability p. Due to symmetry of the normal distribution, the number of equations is not sufficient to solve for both ϵ and σ_w (i.e., $z_p = -z_{1-p}$). There are two possible ways to proceed:

1. Pick a value for ϵ, and then solve for the maximum allowable value of σ_w:

   $$\sigma_w = \frac{\epsilon}{z_{1-p}}$$

   Once data have been gathered and an estimate of σ_w has been computed, the hypothesis:

   $$H_0 : \sigma_w > \frac{\epsilon}{z_{1-p}}$$

   Can be tested against the alternative:

   $$H_1 : \sigma_w \leq \frac{\epsilon}{z_{1-p}}$$

   So, if

   $$\widehat{\sigma}_w \leq \frac{\epsilon}{z_{1-p}} \sqrt{\frac{\chi^2_{df,1-\alpha}}{df}}$$

   where $\widehat{\sigma}_w$ is a sample estimate of σ_w, and $\chi^2_{df,1-\alpha}$ is the $100(1-\alpha)$ percentile of a Chi-squared distribution with df degrees of freedom, then reject H_0 in favor of H_1, and conclude that the measurement system is adequate.
2. Estimate σ_w and then compute the estimated value:

   $$\widehat{\epsilon} = \widehat{\sigma}_w z_{1-p}$$

   If this estimate seems small enough, or to be more conservative, if the upper $100(1-\alpha)$ % confidence limit on ϵ:

$$\widehat{\epsilon}_{1-\alpha} = z_{1-p}\sqrt{\frac{(df)\widehat{\sigma}2}{\chi^2_{df,\alpha}}}$$

seems small enough, where df is the degrees of freedom associated with the estimate $\widehat{\sigma}_w$, then you know your measurement system is adequate.

An Example: No Reference Method Result

A part has a dimension with specification limits $L = 9$, $U = 16$ (the units will go unmentioned). The same five parts will each be measured twice by each of two operators, Operator A and Operator B. The data were entered into a Minitab™ 16 project file. Figure 3.2 shows the data.

The analysis is selected from menu options Stat → Quality Tools → Gage Study → Gage R&R Study (Crossed). Figure 3.3 shows the initial input window.

Figure 3.4 shows the choices made in the "Options" window.

Figure 3.5 shows the output from the Minitab™ session window. There are several references to ANOVA. The reader not familiar with this concept will be afforded more detail about it in further chapters.

The single most valuable result from this output is the Repeatability StdDev (SD), i.e., the standard deviation of measurements within part, within operator. This number is found in the Gage R&R Table, and it is the estimate of $\sigma_w \approx 0.334$. The second most valuable results are the degrees of freedom (column labeled "DF") associated with this "Repeatability" estimate, which, from the ANOVA table, is $df = 10$. We can now make several computations.

$$\frac{6\widehat{\sigma}_w}{U-L} \approx \frac{6(0.334)}{16-9} \approx 0.2863$$

Its reciprocal is:

$$\widehat{C}_p \approx \frac{16-9}{6(0.334)} \approx 3.493$$

Assuming the risk probability $p = 0.001$, and $\alpha = 0.05$,

$$z_{1-p} \approx 3.090$$

$$\widehat{\epsilon} = \widehat{\sigma}_w z_{1-p} \approx 0.334(3.090) \approx 1.032$$

$$\widehat{\epsilon}_{1-0.05} = z_{1-0.001}\sqrt{\frac{(df)\widehat{\sigma}^2_w}{\chi^2_{df,\alpha}}} \approx 3.090\sqrt{\frac{(10)*(0.334)^2}{3.940}} \approx 1.644$$

Thus, we could say that we are 95 % confident that there is no more than a 0.1 % chance that our measurement system would indicate that a part was out of the specification range if its dimension was truly within the interval (9 + 1.644, 16 − 1.644) = (10.644, 14.356).

Session

Worksheet 1 ***

↓	C1-T	C2	C3	C4
	Operator	Part	Rep	measurement
1	A	1	1	11.15
2	A	1	2	10.76
3	A	2	1	12.25
4	A	2	2	12.59
5	A	3	1	13.07
6	A	3	2	12.86
7	A	4	1	13.75
8	A	4	2	13.92
9	A	5	1	14.76
10	A	5	2	15.12
11	B	1	1	10.31
12	B	1	2	10.76
13	B	2	1	12.05
14	B	2	2	12.37
15	B	3	1	12.85
16	B	3	2	12.42
17	B	4	1	13.31
18	B	4	2	14.51
19	B	5	1	14.41
20	B	5	2	15.06
21				

Fig. 3.2 Measurement system example data

When Reference Method Results Are Available

The confidence statement about the risk probability, p, depends upon the assumption that measurement errors are normally distributed, that variation in the parts is also normally distributed, and that the conditional expectation of the measurement, given the true value of the part's dimension, is the true dimension, i.e.

$$E\left(X\middle|Y=y\right) = y$$

In other words, the calculations presume a perfectly accurate, albeit not perfectly precise, measurement system. If, however, the system is not perfectly accurate, then it is possible that

$E(X|Y = y) = f(y)$, where f is some function. Perhaps the most common form of f is linear:

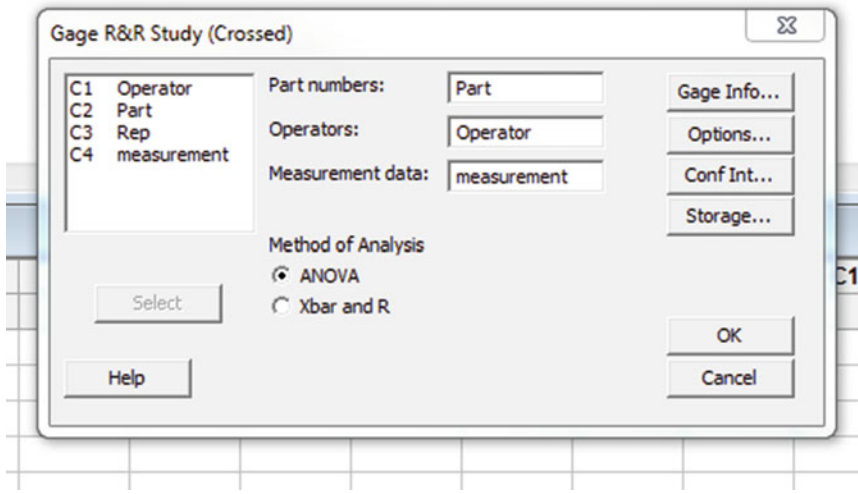


Fig. 3.3 Minitab™ 16 Gage R&R Study (Crossed) initial input window

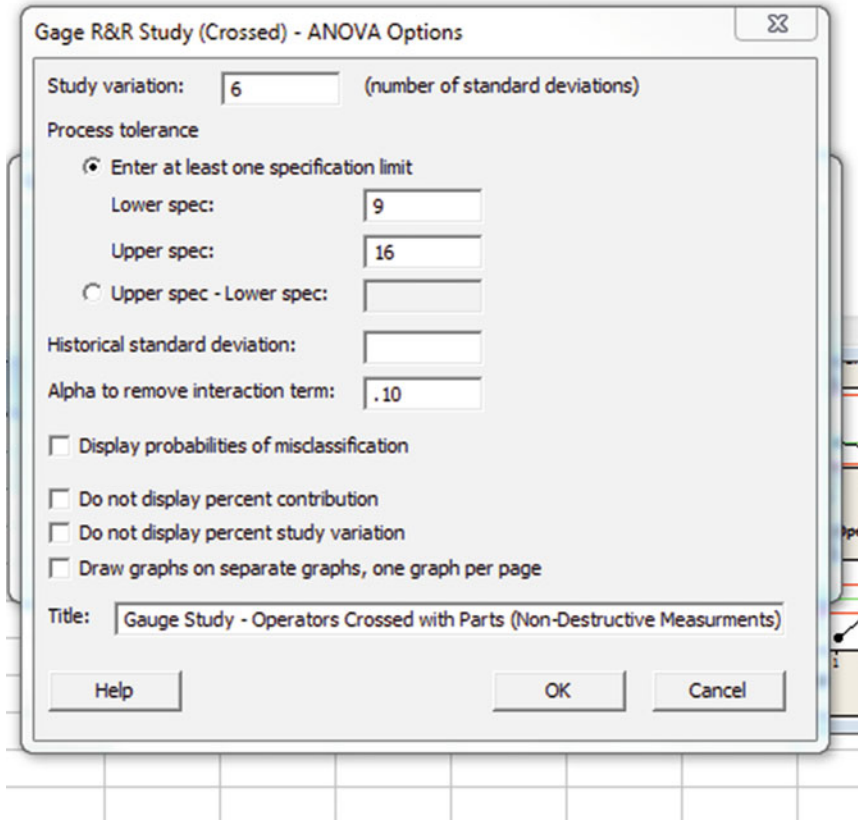


Fig. 3.4 Gage R&R (crossed) options

12/31/2013 12:36:21 PM

Welcome to Minitab, press F1 for help.

Gage R&R Study - ANOVA Method

Gage R&R for measurement

Gage name: Measurement Gauge
Date of study: Now
Reported by: S. Pardo
Tolerance:
Misc:

Two-Way ANOVA Table With Interaction

Source	DF	SS	MS	F	P
Part	4	38.7254	9.68135	272.820	0.000
Operator	1	0.2360	0.23603	6.651	0.061
Part * Operator	4	0.1419	0.03549	0.250	0.903
Repeatability	**10**	1.4200	0.14200		
Total	19	40.5233			

Gage R&R

Source	VarComp	95% CI	%Contribution (of VarComp)	95% CI
Total Gage R&R	0.12401	(0.074, 24.135)	4.93	(0.60, 89.93)
Repeatability	0.11156	(0.060, 0.277)	4.43	(0.36, 15.87)
Reproducibility	0.01245	(0.000, 24.021)	0.49	(0.00, 89.41)
Operator	0.01245	(0.000, 24.021)	0.49	(0.00, 89.41)
Part-To-Part	2.39245	(0.841, 19.955)	95.07	(10.07, 99.40)
Total Variation	2.51646	(0.964, 32.266)	100.00	

Process tolerance = 7

Source	StdDev (SD)	95% CI	Study Var (6 * SD)	95% CI
Total Gage R&R	0.35215	(0.272, 4.913)	2.11291	(1.629, 29.476)
Repeatability	**0.33401**	(0.245, 0.527)	2.00407	(1.467, 3.161)
Reproducibility	0.11157	(0.000, 4.901)	0.66939	(0.000, 29.407)
Operator	0.11157	(0.000, 4.901)	0.66939	(0.000, 29.407)
Part-To-Part	1.54675	(0.917, 4.467)	9.28052	(5.502, 26.803)
Total Variation	1.58633	(0.982, 5.680)	9.51801	(5.892, 34.082)

Source	%Study Var (%SV)	95% CI	%Tolerance (SV/Toler)	95% CI
Total Gage R&R	22.20	(7.73, 94.83)	30.18	(23.27, 421.09)
Repeatability	21.06	(6.02, 39.84)	28.63	(20.96, 45.15)
Reproducibility	7.03	(0.00, 94.55)	9.56	(0.00, 420.10)
Operator	7.03	(0.00, 94.55)	9.56	(0.00, 420.10)
Part-To-Part	97.50	(31.74, 99.70)	132.58	(78.59, 382.90)
Total Variation	100.00		135.97	(84.18, 486.89)

Number of Distinct Categories = 6
95% CI = (0.473271, 18.2322)

Fig. 3.5 Minitab Gage R&R Study (crossed) output

$$E(X|Y = y) = \beta_0 + \beta_1 y$$

The only way to obtain estimates of the parameters β_0 and β_1 would be to have pairs of values (x, y) for at least two parts. That is, for each part, obtain a value x using the measurement system of interest, together with a value y that represents the true dimension. If another measurement method, one that may be more accurate and precise, but possibly more difficult or expensive to use, is available, then a linear regression may be used to estimate the parameters β_0 and β_1.

Suppose the estimates of β_0 and β_1 are b_0 and b_1, respectively. Following the earlier arguments, and generalizing from the conditional probabilities gives:

$$\frac{L - (b_0 + b_1(L + \widehat{\epsilon}))}{\widehat{\sigma}_w} = z_p$$

and

$$\frac{U - (b_0 + b_1(U - \widehat{\epsilon}))}{\widehat{\sigma}_w} = z_{1-p}$$

Solving for $\widehat{\epsilon}$ gives:

$$\widehat{\epsilon}_L = \frac{-z_p\widehat{\sigma}_w + (1 - b_1)L - b_0}{b_1}$$

and

$$\widehat{\epsilon}_U = \frac{z_{1-p}\widehat{\sigma}_w - (1 - b_1)U + b_0}{b_1}$$

Thus, the "point" estimate of the interval of part dimensions for which there is a $100p$ % chance of getting a measurement outside $[L, U]$ is $\left[L + \widehat{\epsilon}_L,\ U - \widehat{\epsilon}_U\right]$.

Unfortunately, obtaining a confidence limit for ϵ_L and ϵ_U is not so simple when $\beta_0 \neq 0$ and $\beta_1 \neq 1$. However, even the point estimates $\left[L + \widehat{\epsilon}_L,\ U - \widehat{\epsilon}_U\right]$ may give enough insight into the adequacy of the measurement system.

Example Revisited: With Reference Method Results

By way of example, suppose in the previous example we had measured the parts using another, much more precise and accurate system. Figure 3.6 shows the original data with the reference method (called "Reference") added.

Figure 3.7 shows a scatter plot of measurement by Reference, with the regression line. Figure 3.8 shows the results of the regression analysis done with Minitab.

With $b_0 = 0.873$ and $b_1 = 0.990$, $p = 0.001$, $\sigma_w = 0.334$, the estimates of ϵ_L and ϵ_U are

$$\widehat{\epsilon}_L \approx 0.252$$

and

Fig. 3.6 Measurement system data with reference added

Worksheet 1 ***

	C1-T	C2	C3	C4	C5
	Operator	Part	Rep	measurement	Reference
1	A	1	1	11.15	10.18
2	A	1	2	10.76	9.91
3	A	2	1	12.25	11.55
4	A	2	2	12.59	11.96
5	A	3	1	13.07	12.41
6	A	3	2	12.86	12.20
7	A	4	1	13.75	12.99
8	A	4	2	13.92	13.26
9	A	5	1	14.76	13.94
10	A	5	2	15.12	14.35
11	B	1	1	10.31	9.47
12	B	1	2	10.76	10.18
13	B	2	1	12.05	11.14
14	B	2	2	12.37	11.60
15	B	3	1	12.85	12.15
16	B	3	2	12.42	11.73
17	B	4	1	13.31	12.58
18	B	4	2	14.51	13.76
19	B	5	1	14.41	13.67
20	B	5	2	15.06	14.19
21					

$$\widehat{c}_U \approx 1.763$$

Thus, the interval of part dimensions that is estimated to yield only a $p = 0.001$ chance of obtaining a measurement outside the specification range is approximately

$$\left[L + \widehat{c}_L,\ U - \widehat{c}_U\right] \approx [9.25,\ 14.24]$$

The approach of evaluating the adequacy of a measurement system described above may be used whenever the measurement of interest is continuously valued, regardless of whether the dimension has a two-sided specification (L and U) or a single-sided specification (either L or U but not both). This is an advantage over the use of C_p, which requires a two-sided specification. Furthermore, the approach can be used whether there does not exist any "reference method" for determining the "true" value of a part's dimension (or performance parameter) or whether such a method exists and a part may be measured both with the system in question and the reference method.

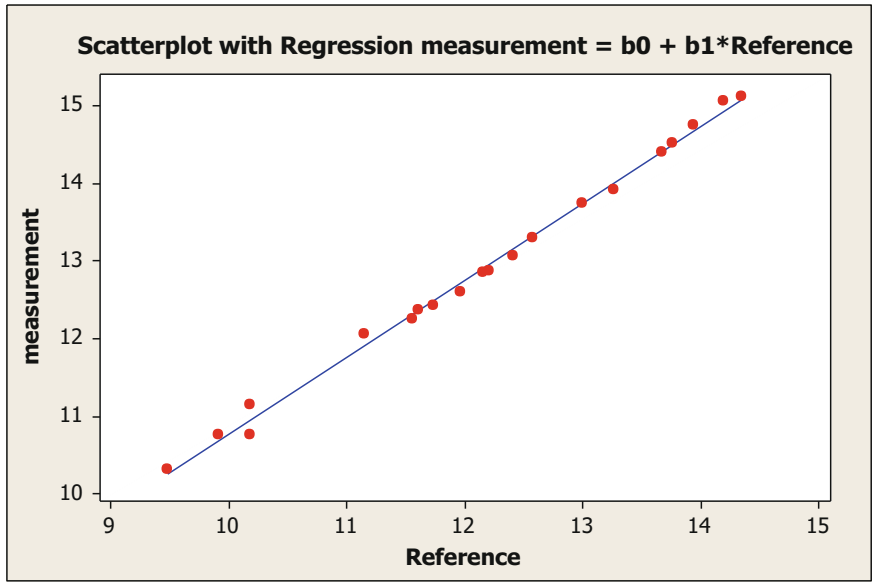


Fig. 3.7 Scatter plot with measurement and reference

Regression Analysis: measurement versus Reference

The regression equation is
measurement = 0.873 + 0.990 Reference

Predictor	Coef	SE Coef	T	P
Constant	0.8734	0.1926	4.53	0.000
Reference	0.99004	0.01573	62.93	0.000

S = 0.100920 R-Sq = 99.5% R-Sq(adj) = 99.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	40.340	40.340	3960.80	0.000
Residual Error	18	0.183	0.010		
Total	19	40.523			

Unusual Observations

Obs	Reference	measurement	Fit	SE Fit	Residual	St Resid
1	10.2	11.1486	10.9520	0.0385	0.1966	2.11R
12	10.2	10.7642	10.9520	0.0385	-0.1879	-2.01R

R denotes an observation with a large standardized residual.

Fig. 3.8 Regression analysis for measurement vs. reference

Concerning Numerical Precision

Data are often reported and recorded based on some number of significant digits (whether or not the data are recorded in scientific or engineering notation). There is no universally accepted standard for how much numerical precision should be used in reporting sample statistics. The topic of numerical precision will be largely ignored in this text. However, this is not to say that it is unimportant. Rather, the EAS must use his or her judgment, based on the particular context. A rule of thumb that some people have employed is that the statistic should be reported to one more decimal place of precision than the measurements themselves. However, even this is not universally true. For example, when reporting proportions, or percentages, the rules guiding the reporting of means and standard deviations may not apply. Furthermore, the numerical precision for reporting p-values may vary from one application/context to another. The experimenter must determine what makes sense in each context.

Key Points

- Experimentation depends on measurement, so insuring that measurement systems are both accurate and precise enough is paramount.
- If a reference measurement is available, it may be possible to assess both precision and accuracy.
- Without a reference, only precision can be assessed.

Exercises and Questions

1. What is required in order to assess accuracy of a measurement system?
2. What is σ_w? Why is it important? How would you estimate it? Can you compute a confidence interval for it?

Chapter 4
Modeling with Data

Engineers and Applied Scientists (EASs) require mathematical models to predict the value of some critical performance variable or some characteristic of a product or process output. Generally, there are two kinds of variables:

1. Inputs
2. Outputs

As a point of terminology, we will often refer to the output variables as "response variables" or simply "responses", and the input variables as "regressors", in that we will rely heavily upon multiple regression methods for building models from empirical observations. Sometimes the input variables are referred to as "factors", especially in relation to a class of data gathering plans called "factorial experiments".

Furthermore, the EAS would like to have a mathematical equation to describe the relationships between the input and output variables, i.e.,

$$\text{Outputs} = f(\text{Inputs})$$

The bad news is that often the function "f" is unknown. The good news is that we have a means of approximating "f", even when it is unknown. The bad news is that we must obtain empirical observations under a variety of conditions in order to approximate f. The good news is that someone has created a means of determining the fewest possible number of conditions under which to gather data in order to approximate f with a polynomial function. The bad news is that in order to accomplish this approximation, we need to know that f is at least piecewise continuous and differentiable. The good news is that most functions we will ever care about are in fact piecewise continuous and differentiable, at least in the range of inputs that is of interest. The bad news is that f may be highly nonlinear, and polynomial functions, even of higher orders, may not "fit" very well. The good news is that we can often make some relatively simple transformations (usually on input variables) so that we can reasonably approximate f as a polynomial in the transformed variables. OK, so we end the discourse about bad news/good news on a good note.

There are some special cases for the function "f" that will be important. The first is the case were all or some of the inputs are discrete, or categorical. In those cases, identical methods will be used to approximate f, but the interpretation will be very different than in the case where the inputs are continuously valued. The other case is where the general form of "f" is known or at least guessed, and it is inherently non-linear in terms of the coefficients or parameters. In this text, we will only consider a very special non-linear case. The reader interested in a more general and complete treatment of non-linear model building should consult Seber and Wild (1989).

S.A. Pardo, *Empirical Modeling and Data Analysis for Engineers and Applied Scientists*,
DOI 10.1007/978-3-319-32768-6_4

Polynomial Approximation

Every student of first year calculus knows that any infinitely differentiable function can be approximated with a Taylor/Maclaurin polynomial, and the error in such an approximation is determined by the order to which one decides to truncate such a polynomial.

In other words, if f is infinitely differentiable, it can be expressed as an infinite Taylor/Maclaurin series:

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k$$

where $f^{(k)}(0)$ is the kth derivative of f evaluated at $x = 0$. Furthermore, we can approximate f with a finite polynomial function of the form:

$$\widetilde{f}(x) = \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k$$

And we can choose n to yield a desired level of error:

$$R(x) = \int_0^x \frac{x^n}{n!} f^{(n+1)}(t) dt$$

Inasmuch as we have no knowledge of f per se, we cannot evaluate any of its derivatives. So how can we evaluate the coefficients of the polynomial $\widetilde{f}(x)$

$$\beta_k = \frac{f^{(k)}(0)}{k!}$$

and how can we possibly evaluate (or control) the error in the approximation?

Empirical Approximation

Let us begin with a motivating, simplified example. Suppose an EAS is designing a joint in which two subassemblies are held together with an epoxy adhesive. Once the epoxy is applied and the two subassemblies are joined, the epoxy must "cure" for some period of time in order to insure that the joint has sufficient strength. Of course, shorter cure times would be desirable, since less time for manufacturing means larger numbers of units produced per unit time. The required minimum pull strength is 7.000 N, or approximately 1.574 lbs. The design problem can be stated as: find the minimum cure time that insures a pull strength no less than 7.000 N.

Inasmuch as the "optimal" cure time is unknown, the EAS performs an experiment in which she makes three pairs of subassemblies, joins them with the epoxy, and allows each to cure. After $\tau_1 = 60.0$ s, she measures the bond's pull strength (force required to pull the two subassemblies apart) of one of the assemblies. After $\tau_2 = 70.0$ s, she measures the pull strength of the second assembly, and after $\tau_3 = 80.0$ s, she measures the pull strength of the third. Her data are shown in Table 4.1.

Table 4.1 Epoxy cure time experiment 1

Time (s)	Pull strength (N)
60.0	3.843
70.0	5.494
80.0	6.710

The EAS makes a plot of her data, shown in Fig. 4.1

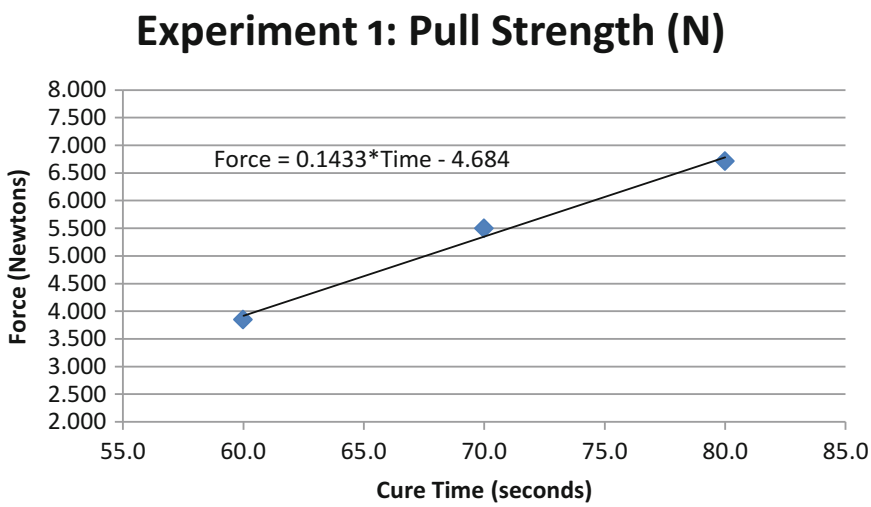


Fig. 4.1 Plot of pull strength data, experiment 1

Table 4.2 Epoxy cure time experiment 2

Time (s)	Pull strength (N)
85.0	6.810
95.0	7.463
105.0	7.273

The EAS noticed that none of her cure times seemed to produce the desired pull strength. She fit a least squares regression (Draper and Smith 1998) line to her data, to obtain a predictive equation, and attempted to project the cure time required to obtain the pull strength of at least 7.000 N. Solving for time, she found that the minimum cure time appeared to be approximately 81.6 s. She runs a second experiment, this time using cure times of 85, 95, and 105 s. Her data for experiment 2 are shown in Table 4.2.

Much to her chagrin, the cure time of 85 s yielded a pull strength of only 6.810 N, and she requires 7.000 N. The cure time of 95 s yielded a higher-than-minimally-desired force (7.463 N). Oddly enough, the pull strength decreased from 7.463 to 7.273 N at a cure time of 105 s. A decrease in pull strength after extending the cure time by 10 s seemed unlikely, but it clearly occurred. Before performing yet a third experiment, she decides to plot all the data from both experiments on a single graph, shown in Fig. 4.2.

Now the EAS has come to three realizations:

1. The function that underlies the relationship between cure time for this epoxy and pull strength is certainly not linear over all time;
2. Either the underlying pull strength function is not monotonically increasing, or there is some noise in the measurements;
3. Extrapolating beyond the range of the observed time points is at best risky.

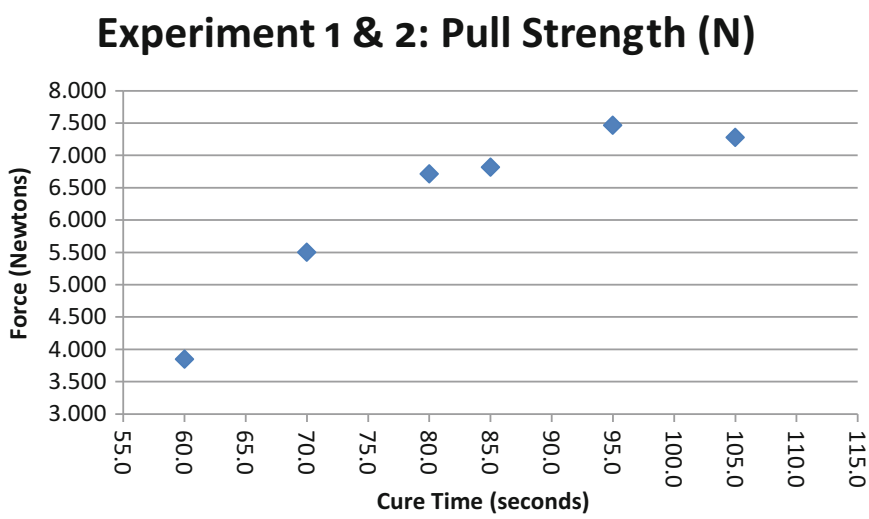


Fig. 4.2 Plot of pull strength data, experiment 1 and 2

Table 4.3 Experiment 1 and 2 with replicates

Experiment	Replicate	Time (s)	Pull strength (N)
1	1	60.0	3.843
1	2	60.0	3.546
1	1	70.0	5.494
1	2	70.0	5.923
1	1	80.0	6.710
1	2	80.0	6.334
2	1	85.0	6.810
2	2	85.0	7.186
2	1	95.0	7.463
2	2	95.0	7.316
2	1	105.0	7.273
2	2	105.0	7.752

The data are plotted in Fig. 4.3

If the function of cure time that describes pull strength were known, and it was differentiable, then the minimization problem would be simple. Differentiate the function with respect to cure time, set the derivative equal to zero, and solve subject to the constraint that pull strength must be at least 7.000 N. In fact, it would be even simpler, in that all the EAS would need to do is find the cure time that yields a pull strength of 7.000 N, as long as the pull strength function is non-decreasing.

Meanwhile, our EAS has decided to repeat the two experiments she performed, so see just how much noise there is in the measurements. She obtained pull strength data for additional subassemblies using the cure times of experiments 1 and 2. Her original data, together with the additional "repeat" data, are shown in Table 4.3. We will refer to the repeated measurements as "replicates".

With these data, the EAS will attempt to approximate the data-generating function using the Taylor series approach. She will first guess at the highest order term to include in her polynomial approximation. Given the apparent curvature (imagine a curve going through the middle of the points in Fig. 4.3) a second order approximation seems appropriate. The approximating function, which we will call the "model" would have the form:

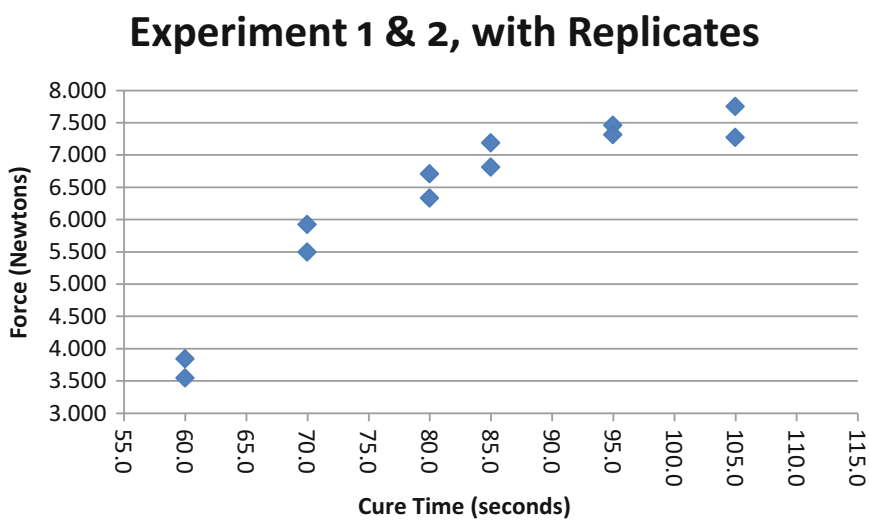


Fig. 4.3 Experiment 1 and 2 with replicated results

$$f(\tau) = \beta_0 + \beta_1\tau + \beta_2\tau^2 + \varepsilon$$

The variable τ represents cure time (in seconds) and the symbol ε represents random "noise", which, for reasons that will become clear later, will be assumed to have a normal (Gaussian) distribution with mean 0 and some standard deviation. The coefficients, β_k, are the Taylor series coefficients:

$$\beta_k = \frac{f^{(k)}(0)}{k!}$$

Using the method of least squares, the data are used to obtain estimates of the β_k (and an estimate of the standard deviation of noise). The predictive model (i.e., the way in which the pull strength is predicted for a given cure time, τ) turned out to be:

$$\widehat{f}(t) = -16.04215 + 0.47377\tau - 0.00238\tau^2$$

Figure 4.4 shows all the data, together with the prediction equation over the cure time range 60–105 s.

Our EAS computed the average result at each of the time points for which she collected data, and then computed, using her empirical model equation, the predicted pull strength at each of those time points. Table 4.4 shows the data with the averages and predictions.

She noticed that the predicted value of pull strength was too low at 80 s and too high at 95 s. Again, using the prediction model, she found that at 84.7 s, the predicted pull strength was approximately 7.002 N (an artifact of rounding time to the nearest tenth of a second), which is only 0.002 N above her specification limit of 7.000 N. So, using data in two separate experiments, with a total of 6 cure times, and two replicates per cure time, the EAS was able to approximate the function that governs how pull strength was related to cure time for this epoxy and the two subassemblies being joined. Furthermore, she was able to use this approximation to determine a cure time that would provide her with the desired pull strength.

There are, unfortunately, some more questions to answer:

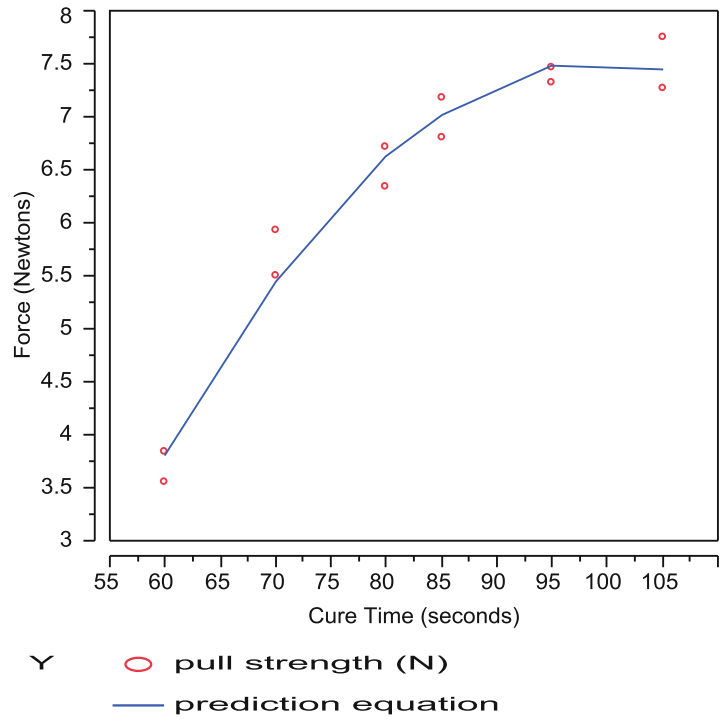


Fig. 4.4 Pull strength data with prediction equation

Table 4.4 Pull strength data with averages, SDs, and predictions

Experiment	Replicate	Time (s)	Pull strength (N)	Ave.	SD	Predicted
1	1	60.0	3.843	3.695	0.2105	3.811
1	2	60.0	3.546			
1	1	70.0	5.494	5.708	0.3035	5.453
1	2	70.0	5.923			
1	1	80.0	6.710	6.522	0.2662	6.618
1	2	80.0	6.334			
2	1	85.0	6.810	6.998	0.2664	7.022
2	2	85.0	7.186			
2	1	95.0	7.463	7.389	0.1033	7.474
2	2	95.0	7.316			
2	1	105.0	7.273	7.513	0.3385	7.448
2	2	105.0	7.752			

1. How can we assess the error in the model? Perhaps we should have carried out the Taylor series to more terms. Perhaps the model is good for predicting the data we observed, but how good would it actually be at predicting the pull strength of a joint cured for 84.7 s?
2. How can we assess the probabilistic variation in future results? If we did the experiment with many subassembly pairs cured at 91.9 s, would all of the resulting joints have a pull strength of 7.385 N? What is the probability that a joint would have a pull strength of at least 7.385 N if cured for 91.9 s?

Examining Model Adequacy

ANOVA (analysis of variance) can be used to decide whether or not the model had any ability to explain or predict the outcomes that were observed. Briefly, ANOVA is a method for partitioning the sums of squared differences between each observation and the mean of all observations, regardless of the conditions under which those observations were obtained. Symbolically, ANOVA provides the terms in the following equation:

$$SST = SSM + SSE$$

SST stands for Total sums of squares, *SSM* stands for Model sums of squares, and *SSE* stands for Error sums of squares. If f_{ij} represents the jth force observed at the ith cure time, and $\overline{f}$ is the average of all the pull strength forces, then

$$SST = \sum_{i=1}^{6}\sum_{j=1}^{2}\left(f_{ij} - \overline{f}\right)^2 = \sum_{i=1}^{6}\sum_{j=1}^{2} f_{ij}^2 - 6 * 2 * \overline{f}^2$$

$$SSM = \sum_{i=1}^{6} + \left(\overline{f}_i - \overline{f}\right)^2$$

$$SSE = \sum_{i=1}^{6}\sum_{j=1}^{2}\left(f_{ij} - \overline{f}\right)^2$$

The analysis of variance (ANOVA) allows us to test whether or not the model accounted for anything more than random variation, completely unrelated to the cure time. In this case, the plot shown in Fig. 4.4 most likely dispels any thoughts that there was no relationship between pull strength and cure time. In more complex situations, where the variable on the left side of the equation is potentially a function of multiple variables on the right, plots may not be so obvious. Therefore, a test of significance is the first line of defense in deciding whether or not a model has any meaning. To perform the test, we compute the ratios:

$$MS_{model} = SSM/(k-1)$$

$$MS_{error} = SSE/(n-1-(k-1))$$

where $n = 6 * 2 = 12$, or the total number of observations used to fit the model, and $k =$ number of parameters (coefficients) in the model $= 3$. *MS* stands for "Mean Square".

Finally, we form the ratio:

$$F = MS_{model}/MS_{error}$$

This ratio, if there is no relationship between the left side of the model equation and the terms on the right (other than noise), and assuming that noise is Gaussian, has an *F* distribution with numerator degrees of freedom $k-1$ and denominator degrees of freedom $n-1-(k-1)$. If this *F* ratio statistic is larger than say the 95th percentile of an *F* distribution with the corresponding degrees of freedom, or analogously if the p-value for the statistic is below a pre-specified value (usually 0.05, or 5 %) then we begin to believe that there is at least some ability of our model to predict, i.e., controlling cure time

Table 4.5 ANOVA table for second order model

Source	DF	Sum of squares	Mean square	F ratio
Model	2	20.4639015	10.2319507	152.613
Error	9	0.60340636	0.06704515	–
Total	11	21.0673078	–	–

has at least some effect on pull strength. Table 4.5 shows the ANOVA results, together with the F ratio. The 95th percentile of an F distribution with 2 and 9 degrees of freedom in numerator and denominator, respectively, is approximately 4.2565. Since the sample F ratio is clearly greater than this, we suspect that the model has some value.

A frequently used measure of goodness for models is called Adjusted R^2. It is referred to as the coefficient of determination (Draper and Smith 1998) and is a measure of the amount of variation in the data explained by the model. It is computed as:

$$R^2_{adj} = 1 - \frac{MSE}{MST}$$

where MST = SST / (n – 1) = Total Mean Square. In this case,

$$R^2_{adj} = 1 - \frac{0.06704}{1.91521} \approx 0.96499$$

The closer Adjusted R^2 is to 1, the "better" the model.

A related quantity, called simply R^2, is given by the formula:

$$R^2 = 1 - \frac{SSE}{SST}$$

Inasmuch as this quantity does not include the degrees of freedom, it does not account for the sample size or the number of parameters in the model. Therefore, we will avoid using it or discussing it any further.

An important assumption required to make the F ratio actually have an F distribution is that the random errors, or noise, associated with observations, are not only normally distributed, but the noise has the same standard deviation, or variance, regardless of the values of the regressors. The condition of constant variance is the most important assumption. Violation of this assumption is referred to as heteroscedasticity (Armitage 1971).

Another question is, why stop at a second-order model? The EAS has a total of 6 cure times at which the pull strength was measured. She could fit up to a fifth order polynomial to her data. Surely the Taylor series approximation would be improved with every additional term. However, she would potentially be making the mistake of over-parameterizing. That is, suppose that in truth the underlying function was a second-order polynomial. Then adding three other terms would only be adding noise to the fit. In fact, adding too many terms might make the model fit quite well, but it would have poor predictability. There is a balance between reducing errors in the model with respect to actual observations and making a model that is perfect for the observed data but predicts future observations poorly. The balance is chosen to satisfy the dictum (razor) of William Ockham (or sometimes spelled Occam), circa 1320 CE:

Pluralitas non est ponenda sine neccesitate

Or, roughly, don't over-explain the data (but don't under-explain them either). The rule of Ockham's Razor is also referred to as the rule of parsimony.

Table 4.6 C_p calculations

n	Order	p	MSE_{max}	SSE_p	C_p	Adj. R^2
12	Fifth	6	0.06726	0.40357	6.00	0.96488
12	Fourth	5	0.06726	0.43267	4.43	0.96773
12	Third	4	0.06726	0.47941	3.13	0.96871
12	Second	3	0.06726	0.60341	2.97	0.96499
12	First	2	0.06726	3.59026	45.38	0.81254

One method for choosing the "right" order model employs Mallow's C_p (Draper and Smith 1998). This C_p statistic is not to be confused with the capability index C_p as it applies to measurement systems analysis. Here is how Mallow's C_p works:

The highest order model the EAS can fit to these data is fifth order. A fifth order polynomial would thus have six parameters (including the intercept). Suppose she fits this model and obtains the *SSE* and MSE for it. Call the *MSE* for the fifth order model MSE_{max}. Then she proceeds to fit fourth, third, second, and first order models, and records the MSE (not the SSE) for each. Let SSE_p represent the *SSE* for a model with p parameters. Mallow's C_p statistic is:

$$C_p = \frac{SSE_p}{MSE_{max}} - (n - 2p)$$

where n = the total sample size (in this case $n = 12$), and p = number of parameters in the model (including the intercept). The idea is that if the (p-1)th order model was correct, then the expected value of C_p would be approximately:

$$E[C_p] = E\left[\frac{SSE_p}{MSE_{max}}\right] - (n - 2p) \approx \frac{(n-p)\sigma^2}{\sigma^2} - (n - 2p) = p$$

So, the idea is to compute C_p for all models up to max(p) – 1, and choose the model with the smallest C_p that is approximately equal to p. Table 4.6 shows all the computations for C_p and for Adjusted R^2.

Thus, the second order model (with $p = 3$) has the smallest C_p that is approximately equal to p. Although the second-order model does not have the highest Adjusted R^2 of the models fit to the data, it is probably the best choice, since it not only has the lowest value of C_p. But also its C_p value is very close to $p = 3$. The EAS concludes that the best polynomial approximation to the pull strength as a function of cure time is second order (quadratic).

Examining Variation

The residuals, or differences between actual observations and predicted values, reveal a lot about variability. Figure 4.5 shows a histogram of the residuals for the second order model.

With only $n = 12$ residuals, the histogram is not a good indicator of whether the residuals are distributed normally (Gaussian). However, the standard deviation of the residuals gives some indication of how much noise may be affecting the measurements. With a residual standard deviation of $s \approx 0.2342$ N, we can get some idea of how much variation there would be in a predicted value of pull strength at any given cure time. Another estimate of noise standard deviation is obtained from the ANOVA for the model. The ANOVA algebraically "partitions the total variation of the data into two additive terms, one for the "model" and one for error, or noise. From this partitioning, we obtain another estimate of the noise standard deviation, sometimes called root mean square error, or RMSE. For these data with the quadratic model the EAS fit, the RMSE was $s \approx 0.2589$.

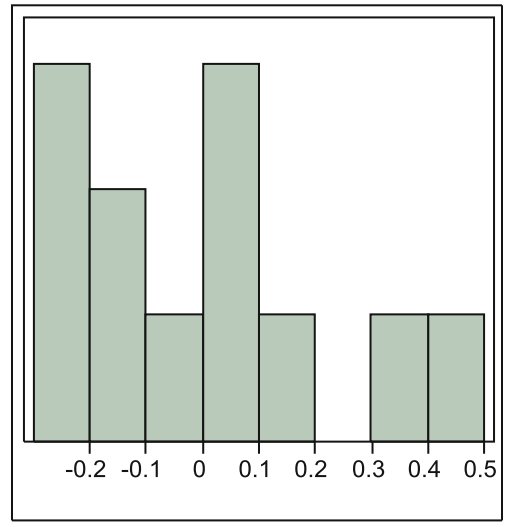


Fig. 4.5 Histogram and sample statistics for residuals from second order model

Quantiles

100.0%	maximum	0.4704
99.5%		0.4704
97.5%		0.4704
90.0%		0.4204
75.0%	quartile	0.1461
50.0%	median	0.0110
25.0%	quartile	-0.2032
10.0%		-0.2786
2.5%		-0.2844
0.5%		-0.2844
0.0%	minimum	-0.2844

Moments

Mean	-7.4e-17
Std Dev	0.2342117
Std Err Mean	0.0676111
Upper 95% Mean	0.148811
Lower 95% Mean	-0.148811
N	12

The standard error of a prediction (assuming that the model is actually perfect, which is rarely actually true) is given by the formula:

$$SE\left(\hat{y}\middle|\boldsymbol{\tau}_0\right) = s\sqrt{\boldsymbol{\tau}_0'\left[\boldsymbol{T}'\boldsymbol{T}\right]^{-1}\boldsymbol{\tau}_0}$$

where:

$$\boldsymbol{\tau}_0 = \begin{bmatrix} 1 \\ \tau_0 \\ \tau_0^2 \end{bmatrix}$$

and:

$$\boldsymbol{T} = \begin{bmatrix} 1 & \tau_1 & \tau_1^2 \\ \vdots & \vdots & \vdots \\ 1 & \tau_{12} & \tau_{12}^2 \end{bmatrix}$$

The $'$ indicates matrix transpose.

The variable τ_0 represents the cure time of interest. With $n = 12$, and $\tau_0 = 84.7$,

$$SE\left(\widehat{y}\middle|\tau_0 = 84.7\right) \approx 0.1085$$

What are the implications? The EAS can use this standard error calculation to create a 95 % lower confidence limit for a predicted value at 84.7 s. That is, the value:

$$\widehat{y} - t(0.95, n-p)SE\left(\widehat{y}\middle|\tau_0\right)$$

is a feasible lower limit for what to expect as a predicted pull strength at τ_0 cure time seconds. The letter t represents a percentile from a Student's t distribution. In this particular case,

$$t(0.95, 12-3=9) \approx 1.8331$$

The 95 % lower confidence limit for a predicted value of pull strength at 84.7 s is therefore approximately 7.002 – 1.8331 * 0.1085 ≈ 6.803 N. In other words, if the model is correct, then there is approximately a 95 % chance that the predicted value of pull strength at 84.7 s would be no less than 6.803 N.

Some people may be tempted to interpret the limit as a 95 % percentile of pull strengths at 84.7 s. This is not correct. The limit gives 95 % confidence about predicted values coming from this quadratic model with $n = 12$ observations. The limit is a measure of repeatability of predictions. Given that the lower specification limit for pull strength is 7.000 N, and given that the lower confidence bound on predicted values at 84.7 is 6.803 N $<$ 7.000 N, the EAS is concerned that 84.7 may not be adequate cure time. After all, it is feasible that she could have gathered a different set of subassemblies, joined them and measured pull forces at the same cure times as in her experiments 1 and 2, fit a second order polynomial model, and obtained a predicted pull strength value less than 7.000 N. The EAS wants to have a high level of confidence (99 %) that the pull strength of joints will be at least 7.000 N. In other words, she wants to find the smallest value of τ_0 such that:

$$\widehat{y} - t(0.99, n-p)SE\left(\widehat{y}\middle|\tau_0\right) \geq 7.000$$

Using a numerical solver program, she finds that a cure time of 89.7 s yields a lower 99 % confidence limit for predicted value of pull strength to be approximately 6.988 N.

Verification

Once our EAS has made the decision to set the cure time to 89.7 s, she decides wisely to validate her decision. She looks at the data she had already gathered, and notices that only at 95 s and 105 s were all observations of pull strength over the 7.000 N limit. She computes the 99 % lower confidence limits of predicted pull strength at 90, 92.5, and 95 s, which are 7.0007 N, 7.100 N, and 7.160 N,

Table 4.7 Validation experiment results

Time (s)	Pull strength (N)	Mean pull	SD pull
89.7	7.581	7.305	0.3904
89.7	7.029		
92.5	7.320	7.203	0.1644
92.5	7.087		

Table 4.8 ANOVA for second order model with additional data

Source	DF	Sum of squares	Mean square	F ratio
Model	2	23.2476748	11.6238374	252.2918
Error	13	0.59894883	0.04607299	
C. total	15	23.8466237	–	–

Table 4.9 Mallow's C_p and adjusted R^2 original and with new data

n	Order	p	MSE_{max}	SSE_p	C_p	Adj. R^2
12	Fifth	6	0.06726	0.40357	6.00	0.9649
12	Fourth	5	0.06726	0.43267	4.43	0.9677
12	Third	4	0.06726	0.47941	3.13	0.9687
12	Second	3	0.06726	0.60341	2.97	0.9650
12	First	2	0.06726	3.59026	45.38	0.8125
16	Second	3	0.06773	0.8506	2.56	0.9590
16	Third	4	0.06773	0.85060	4.56	0.9637
16	Fourth	5	0.06773	0.66551	3.83	0.9621
16	Fifth	6	0.06773	0.62702	5.26	0.9608
16	Sixth	7	0.06773	0.60959	7.00	0.9576

respectively. Being cautious, she decides to repeat the experiment at 89.7 s, and at 92.5 s (half-way between 90 and 95 s). Her results are given in Table 4.7.

Based on these results, it appears 89.7 s is sufficient. However, standard deviation of pull strength at 89.7 s is approximately 0.3904, and 0.1644 at 92.5 s. Of course, these are all computed using only $n = 2$ observations. Although the average results at both 89.7 and 92.5 exceeded the specification limit, the EAS is concerned about individual results due to the relatively large standard deviations. Noticing that all other conditions (raw material lots, machinery, even operators) have not changed between her model-fitting runs and the "verification" runs, the EAS decides to add the new data to the old and refit the second order model. The new model is:

$$\widehat{f}(t) = -16.08556 + 0.47297\tau - 0.00238\tau^2$$

The adjusted R^2 for this model is 0.9364, which is lower than the model with the earlier data. The ANOVA for the new model is given in Table 4.8.

The critical value for the F Ratio is approximately 3.8056, so clearly the model has meaning. Relative to a fifth order model with the new data, Mallow's C_p is 8.86, which is much higher than the second order model with the original data. The EAS decides to fit third, fourth, fifth, and sixth order models with the new data. Table 4.9 shows the C_p calculations, together with the adjusted R^2 values.

The fifth order model has a C_p that is closest to the theoretical ideal ($C_p \approx p$), and it has a fairly high adjusted R^2. Therefore, the EAS decides to use the fifth order model fit to the total dataset. She then computed the lower bound on the predicted pull strength at each time point she observed. Figure 4.6 shows the observed data, the predicted values and the 99 % lower bounds. Drawing an horizontal line at pull strength = 7.0, and noting where it intersects the 99 % lower limit, the EAS decides that her best choice of cure time is approximately 92.5 s.

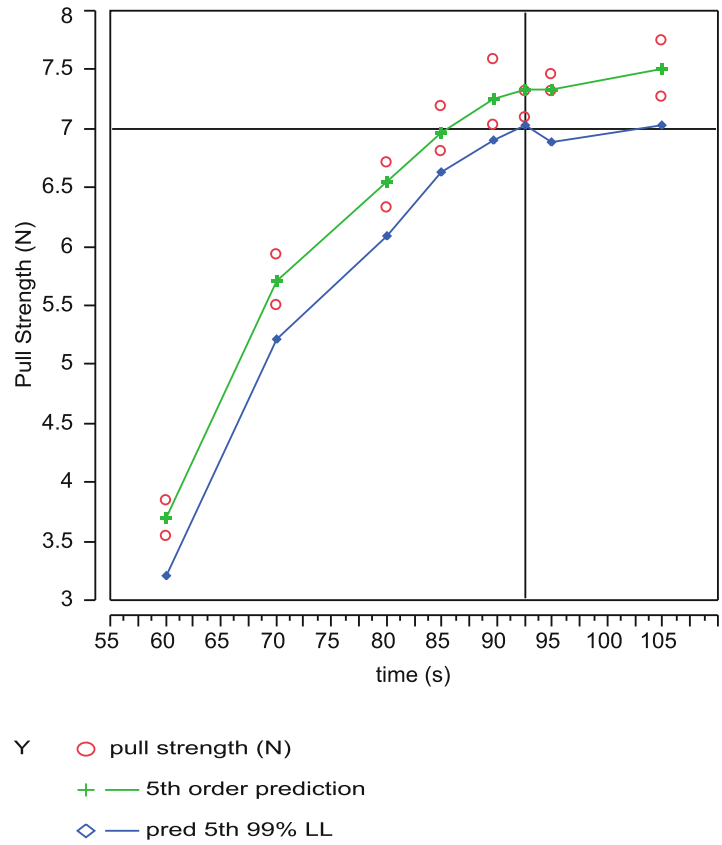


Fig. 4.6 All data with fifth order model predictions and prediction limits for pull strength

There is one small issue. Our EAS reads a book on regression methods, such as Draper and Smith (1998), and discovers that the standard error formula she used to compute the lower 99 % bound on predicted values did not account for variation in future values of observations. To account for the additional variability, she would have to use the standard error formula:

$$SE\left(\widehat{y}\middle|\boldsymbol{\tau}_0, future\right) = s\sqrt{1+\boldsymbol{\tau}_0'\left[\boldsymbol{T}'\boldsymbol{T}\right]^{-1}\boldsymbol{\tau}_0}$$

She also discovers another type of limit, called a precision limit (Dunn 2010) that is computed as:

$$\widehat{y} - t(0.99, n-p)s$$

where s is the root mean square error from the regression. She plots all the limits, together with the predicted values, over the entire range of times used in her experiments. The plot is shown in Fig. 4.7.

Now the EAS is not so certain that 92.5 s is in fact adequate. The most optimistic limit, 99 % prediction lower limits, indicates that 92.5 s would be adequate with 99 % probability. The most pessimistic limit, future value 99 % lower limits, and even the next most pessimistic limit, the precision 99 % lower limit, indicate that 105 s may not be adequate. The model indicates an inexplicable dip in pull strength between 92.5 and 95 s. All the observations past 92.5 s (admittedly only four values) gave pull strengths over the lower specification limit of 7.000 N. The EAS decides that the dip in pull strength may be due to random effects, and the model is falsely indicating lower than desirable pull strengths between 92.5 and 105 s. So, she decides to hedge her bets, and set the time to 95 s. However, she is not so careless as to run at 95 s without at least a small test.

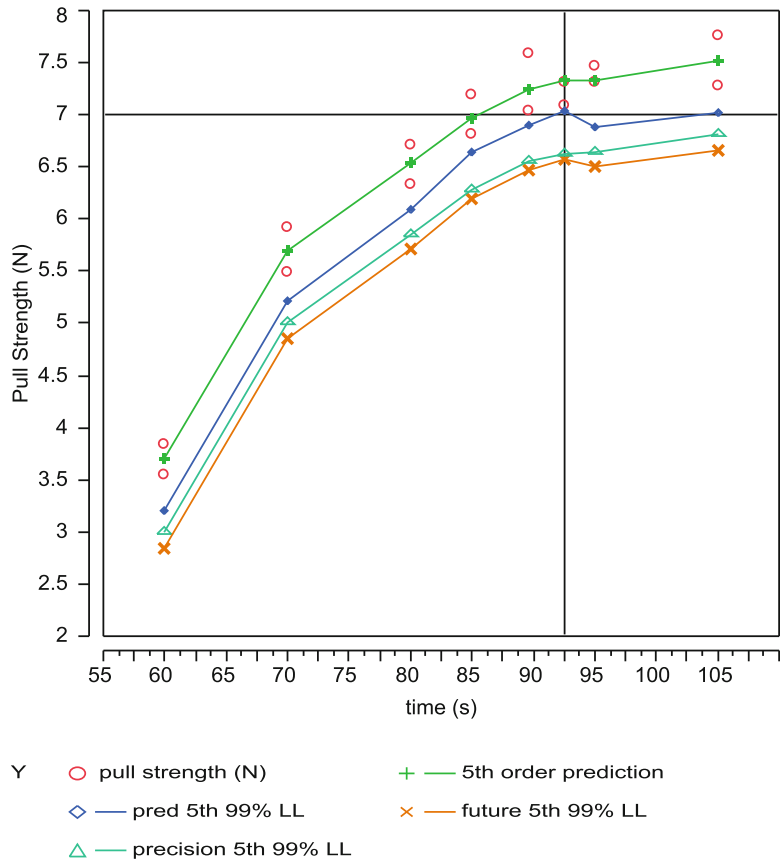


Fig. 4.7 Pull strength data with fifth order model predictions and 99 % lower limits

Table 4.10 Confirmatory tests at cure time = 95.0 s

Time (s)	Pull strength (N)	Mean pull	SD pull
95.0	7.639	7.466	0.2455
95.0	7.292		

To confirm her decision, the EAS makes two more tests at a cure time of 95.0 s. The results are shown in Table 4.10. It appears that her decision was well-founded, as the two pull strengths were both greater than 7.0. Furthermore, the standard deviation is low enough so that the specification limit approximately 1.9 standard deviation units below the sample mean of 7.466. The sample z-score for the limit of 7.000, based on the mean and standard deviation of these two values, yields approximately a 97 % chance of obtaining pull strengths above the lower limit. Therefore, the EAS has strong confidence that the cure time of 95.0 s will in fact yield acceptable pull strengths.

What We Have Discovered

Sometimes we are trying to make a decision about one controllable variable or set of controllable variables that affect a critical output or response variable (or set of variables). It may be that there is some functional relationship between the input and output variables, but its exact mathematical form may be unknown. We presumed that this unknown mathematical form is “smooth”, or in other words,

infinitely differentiable. Thus we employed a Taylor series sort of argument as an attempt to approximate the function with a polynomial.

Choosing the order of the polynomial approximation is not a simple process. Even in the case of a single input variable (e.g., cure time) we were not certain what order to choose. If we had only obtained functional values (outputs, i.e., pull strength) at two different values of the input variable, we could only have approximated the relationship with a first order polynomial. Recognizing the possibility that a first order approximation might be insufficient, the EAS chose three points. There is no absolute rule about the order of approximation; choosing is more art than science. Clearly the number of points should exceed the highest order approximation you would ever want to consider by 1. So, if you thought that the highest order approximation you would want was second order, then select three input values. Often the choice of approximation is constrained by resources; you may only have budget or time enough to collect data at two input variable points. The EAS must make stakeholders and decision-makers aware of the limitations and risks associated with the experimentation. Fewer input points results in a cruder approximation. In the case of the example, not only was the linear approximation inadequate, but the range of input points did not include the "optimal" point. It may be possible to make decisions about experimental conditions in a sequential fashion, adding new data to previously collected data. Once the EAS discovered that her three input point did not seem to cover the desired output, she was able to add data from additional points. In her case, interfering conditions such as raw material lots, different machines/production lines, operators, and test equipment were not an issue. Thus, she had the luxury of adding points as she performed analyses.

We saw that the method of least squares can be used to find an approximating polynomial. Inasmuch as the approximation is made using empirical observations, and that observations have a random component, some statistical methods should be employed to determine whether the approximation was reasonable. The first tool was the ANOVA for the model, which indicates whether or not there is any relationship between the approximation and the actual underlying function. We used the F-ratio test and the adjusted R^2 statistic to decide if our approximating process was on track. We then used Mallow's C_p to help choose an order of approximation. By obtaining data at a sufficient number of input points, we were able to compare C_p for a number of polynomials, and we were able to choose the polynomial approximation for which C_p was closest to the number of parameters in the model. We also want adjusted R^2 told be as close to 1 as possible, but sometimes we may decide to trade increased R^2 for a closer to optimal C_p. Using the C_p criterion is in concert with Ockham's Razor, or the principle of parsimony in model building.

Once the model is built, the next step is to determine the optimal input point, that is, the point at which the output is closest to whatever value we desire. Desires are usually either to minimize, maximize, obtain a particular threshold value (either minimum or maximum) or obtain a specific target value for the output. If you want to maximize or minimize some objective function "on the average", it may be sufficient to find an adequate polynomial model fit to data, and then use deterministic optimization methods (e.g., modified Newton-Raphson methods) to find the optimal point, at which no data may have been gathered yet. In any empirical investigation, the variation in sample results should be considered. Once an optimal point is selected, some additional data at (or at least near) the optimal point should be obtained. These data may verify that the selected point is in fact acceptable, or they might indicate that there is enough variability in the response to warrant accounting for a margin of error. Confidence limits for predicted values, or prediction limits, may be useful in choosing an operating point that provides adequate margin.

We are not necessarily advocating incrementally increasing the range of the experimental factor (s), in this case cure time, and consequently incrementing the order of polynomial approximation. More frequently, there is more than one factor involved, and choosing experimental conditions in a multifactor experiment is better achieved through the use of factorial experimental design, the topic

of the next chapter. In this chapter we are trying to illustrate how an unknown function can be reasonably approximated by a polynomial fit to data.

A Note About Outliers

Sometimes a small number of points can be so influential as to drastically alter the order of the "best" approximating polynomial. Such points are often referred to as "outliers". Be aware that just because a point is highly influential does not make it "wrong", or worthy of discarding. Much more could be said about "outliers". For now, it is sufficient to be aware that there are some diagnostic calculations for identifying points as highly influential, and that identification of high influence is not sufficient reason to discard a point from analyses or model-building.

Key Points

- Models are useful for approximating real phenomena.
- Polynomial models can be built using empirical observations via least squares regression.
- Models can be assessed for adequacy.
- Adjusted R^2 is one measure of model adequacy; ANOVA and confidence intervals for predicted values are other measures.
- The order of the polynomial fit should follow Ockham's Razor, or the rule of parsimony.

Exercises and Questions

1. Do you agree with the decision our EAS made about the pull strength problem? Would you have done anything differently?
2. What are the issues associated with over-fitting, or having too many parameters/terms in a model?

Chapter 5
Factorial Experiments

We have emphasized the need of the EAS to construct an approximating function to relate product design features to performance measures. The EAS needs a method for choosing the different combinations of input feature/characteristic values in the most efficient manner possible. Also, sometimes the EAS is faced with the problem of deciding which smaller subset of too many input variables are most important, that is, have the greatest influence on the response. Attempting to optimize a response over many inputs may be at best difficult, if not completely impractical. The EAS will need a plan that involves the fewest number of input variable points to determine whether or not each potential input variable should or should not be investigated further. This chapter will be largely concerned with making such plans, which are termed "factorial experiments". In this context, the input variables will often be referred to as "factors".

The initial screening of factors involves a linear (first order) approximation to the response function. For each input variable, the question to be answered is whether changing the variable increases or decreases the average value of the response, or if changing the input variable induces no change on the average. Admittedly, this assessment is crude. However, if the EAS is faced with more input variables that can be simultaneously optimized in an economically feasible fashion, it may be the best assessment that can be made. Suppose, for example, there were five input variables that were of potential interest. In order to fit a fourth order polynomial in each input variable simultaneously, the EAS would require $5^5 = 3125$ points. Imagine if the EAS wanted to duplicate results at each point. The EAS would require $N = 6250$ measurements of the response variable. This is probably infeasible. If, however, the EAS desires to decide whether any of the input variables could be disregarded initially, he should consider using a linear approximation to the response function.

In order to fit a linear function to data, two points are minimally required for each input variable. The potential values to be included in the experiment for a given factor are called "levels". In the case of five input variables, the number points would be $2^5 = 32$, corresponding to all possible combinations of two levels per factor, with $r = 5$ factors. This is considerably fewer than 3125.

A first order model may have the form:

$$y = \beta_0 + \sum_{i=1}^{r} \beta_i x_i + \sum_{i \neq j} \sum \gamma_{ij} x_i x_j + \varepsilon$$

The symbol ε represents the random "noise" that is associated with response values observed or measured under "identical" conditions, at least in terms of the input variables. The EAS could include higher-order cross-product terms in the first order model, e.g.:

S.A. Pardo, *Empirical Modeling and Data Analysis for Engineers and Applied Scientists*,
DOI 10.1007/978-3-319-32768-6_5

$$x_i x_j x_k (3\text{-way})$$
$$x_i x_j x_k x_l (4\text{-way})$$
$$\vdots$$
$$x_1 x_2 x_3 \ldots x_r (r\text{-way})$$

Generally, for a first order model used primarily to decide which inputs are truly important, anything beyond the two-way cross-products is overkill.

In order to obtain statistical estimates of the coefficients β_i and γ_{ij}, some minimum number of points must be selected for data gathering. Suppose there were k input variables, and the EAS wanted to estimate coefficients for a "complete" linear model, i.e., including all possible cross-product terms. Then, as mentioned earlier, the minimum number of points would be 2^k. So, if $k = 5$, $2^5 = 32$ points would be required. With duplicate response values at each point, a total of $n = 64$ measurements would be required.

Assessing the Effect of Each Factor

In the case of a single input variable, or factor, to assess the linear effect of the factor, x, on a response variable, y, we require observations of y at two different values, or levels, of x. The levels chosen should cover the range of interest. That is, the lower of the two levels should be the lowest value for which x could feasibly be set, and the highest level should be the highest value at which x could feasibly be set. Denote the low and high levels of x by x^- (for low) and x^+ (for high). Let y^- represent the value of response y^+ observed at x^-, and y^+ the value of the response at x^+. The effect of x is estimated by

$$E = y^+ - y^-$$

If n observations or measurements of the response variable are obtained at each condition, then the average value of the response under each condition is used to compute the effect:

$$E = \overline{y}^+ - \overline{y}^-$$

Notice that the slope of the line segment joining the points, (x^-,y^-) and (x^+,y^+) is given by:

$$b = \frac{\overline{y}^+ - \overline{y}^-}{x^+ - x^-} = \frac{E}{x^+ - x^-}$$

Now suppose we computed the midpoint between x^- and x^+:

$$m = \frac{x^- + x^+}{2}$$

Then transform, or code, the x variable into a new variable whose range is $(-1, +1)$ with midpoint 0:

$$H = \frac{x - m}{\frac{1}{2}(x^+ - x^-)}$$

When $x = x^-$, $H = -1$; when $x = x^+$, $H = +1$; when $x = m$, $H = 0$. Now the new slope, call it b', with respect to the coded input variable, w, is given by:

$$b' = \frac{(\overline{y}^+ - \overline{y}^-)}{2} = \frac{E}{2}$$

This coding transformation is called Helmert coding, named after Professor Dr. Friedrich Robert Helmert. Unless otherwise specified, we will assume that all input variables are expressed as Helmert-coded. Thus -1 will represent the lowest level, and $+1$ the highest level, of each factor in the experimental design.

We would like to gather data under various conditions determined by multiple input factors in such a way as to allow the independent estimation of all coefficients in the first order model. Each unique condition will be called a "run". The collection of conditions used to gather the data for fitting the model (i.e., estimating the coefficients) is called an experiment. The collection of conditions is also called an experimental design, or simply "design".

Assessing the Cross-Product, or Interaction Effects

Suppose there is more than one input factor that may have some effect on the response. It is possible that the first order approximation model should include at least the two-way cross-product terms. Cross-product terms are also called *interaction effects*, in that the levels of all factors included in the cross-product "interact" in their joint effect on the response. Just as in the main effect of each factor individually, and interaction effect can be computed. The two-way interaction effect for any two factors would be the difference between the effect of the first factor when the second is at its "low" (-1) level, and the effect of the first factor when the second is at its "high" $(+1)$ level. The choice of which factor is "first" and which is "second" is arbitrary. If there were exactly two input factors, x_1 and x_2, coded to $(-1, +1)$, the experimental runs could be symbolized as in Table 5.1.

If the symbols y(−,−), y(−,+), y(+,−), and y(+,+) represent the average responses in each run, then the effect of $x1$ at each level of $x2$ can be computed as:

$$E_{x2=-1} = y(+,-) - y(-,-)$$

$$E_{x2=+1} = y(+,+) - y(-,+)$$

The interaction effect is computed as:

$$E_{x1,x2} = \frac{E_{x2=+1} - E_{x2=-1}}{2}$$

Table 5.1 A two-factor, two-level experiment

Run	x_1	x_2	y
1	−1	−1	y (−,−)
2	−1	+1	y (−,+)
3	+1	−1	y (+,−)
4	+1	+1	y (+,+)

The coefficient corresponding to this cross-product term is:

$$b_{12} = \frac{E_{x1,x2}}{2}$$

It turns out that the coefficients computed in the fashion described are actually the least squares estimates.

A Three-Factor Example

Consider the following example. The processing of a liquid chemical mixture can have different yields (in volume), depending on how the temperature (°C), pressure (kP), and dwell time (s) are set. Nominally, the temperature has been set at 37°C, the pressure at 12.00 kP, and dwell time of 300 s. The yield has been averaging 20 liters (L) per batch. An EAS is assigned the task of increasing the volume yield, if possible. He decides to perform an experiment to fit a first order model in the three factors. The levels and their coded values for all the runs are shown in Table 5.2.

There are $2^3 = 8$ runs, representing all possible combinations of levels for the three factors. The model to be fit is:

$$y = \beta_0 + \beta_1 H_1 + \beta_2 H_2 + \beta_3 H_3 + \gamma_{12} H_1 H_2 + \gamma_{23} H_2 H_3 + \gamma_{13} H_1 H_3 + \delta_{123} H_1 H_2 H_3 + \varepsilon$$

The three-way cross-product term is probably not necessary, but since all possible combinations of factor levels are included in the design, it is possible to estimate the coefficient for this term. The EAS decides to get $n = 2$ duplicate values for each run. The data are given in Table 5.3.

There are two very important properties of this experimental design that are made apparent by the use of Helmert coding. The first is balance; for any input factor, there are equal numbers of observations made when the input factor is set to its low (−1) value and its high (+1) value. If you sum all the −1's and +1's in any column, the result is zero. The second is orthogonality; if each of the columns in Table 5.3 are thought of as column vectors, the dot product of any two columns is zero. Balance and orthogonality greatly simplify the calculations of least squares.

Once the data are gathered, the analysis begins with a least squares fit, or estimation of the coefficients in the first order model.

With two-level experiments, where each input variable is Helmert-coded, there is a simple way to compute the effects and coefficients for each term in the first order model. Suppose h_{ij} represents the jth coded value for the ith input variable, and y_j represents the jth value of the response variable. If there are $m = 2^k$ unique runs in the experiment, with n replicate values for each run, then the estimate of the effect, E_i, and the least squares estimate of the coefficient, β_i, are given by:

Table 5.2 Experimental design—volume yield

	Natural			Coded		
Run	Temp	Pressure	Time	H_1	H_2	H_3
1	32	10	240	−1	−1	−1
2	32	10	360	−1	−1	+1
3	32	14	240	−1	+1	−1
4	32	14	360	−1	+1	+1
5	42	10	240	+1	−1	−1
6	42	10	360	+1	−1	+1
7	42	14	240	+1	+1	−1
8	42	14	360	+1	+1	+1

Table 5.3 Volume yield experiment

Run	Temp	Pressure	Time	H_1	H_2	H_3	Volume
1	32	10	240	−1	−1	−1	20.8
1	32	10	240	−1	−1	−1	20.8
2	32	10	360	−1	−1	1	21.6
2	32	10	360	−1	−1	1	21.3
3	32	14	240	−1	1	−1	17.8
3	32	14	240	−1	1	−1	17.7
4	32	14	360	−1	1	1	17.9
4	32	14	360	−1	1	1	17.7
5	42	10	240	1	−1	−1	17.6
5	42	10	240	1	−1	−1	17.8
6	42	10	360	1	−1	1	16.5
6	42	10	360	1	−1	1	16.7
7	42	14	240	1	1	−1	24.1
7	42	14	240	1	1	−1	24.4
8	42	14	360	1	1	1	23.3
8	42	14	360	1	1	1	23.0

Table 5.4 Volume yield data with calculations of some of the coefficient estimates

Run	Temp	Pressure	Time	H_1	H_2	H_3	y = volume	H_1*y	H_2*y	H_3*y	H_1*H_2*y
1	32	10	240	−1	−1	−1	20.8	−20.8	−20.8	−20.8	20.8
1	32	10	240	−1	−1	−1	20.8	−20.8	−20.8	−20.8	20.8
2	32	10	360	−1	−1	1	21.6	−21.6	−21.6	21.6	21.6
2	32	10	360	−1	−1	1	21.3	−21.3	−21.3	21.3	21.3
3	32	14	240	−1	1	−1	17.8	−17.8	17.8	−17.8	−17.8
3	32	14	240	−1	1	−1	17.7	−17.7	17.7	−17.7	−17.7
4	32	14	360	−1	1	1	17.9	−17.9	17.9	17.9	−17.9
4	32	14	360	−1	1	1	17.7	−17.7	17.7	17.7	−17.7
5	42	10	240	1	−1	−1	17.6	17.6	−17.6	−17.6	−17.6
5	42	10	240	1	−1	−1	17.8	17.8	−17.8	−17.8	−17.8
6	42	10	360	1	−1	1	16.5	16.5	−16.5	16.5	−16.5
6	42	10	360	1	−1	1	16.7	16.7	−16.7	16.7	−16.7
7	42	14	240	1	1	−1	24.1	24.1	24.1	−24.1	24.1
7	42	14	240	1	1	−1	24.4	24.4	24.4	−24.4	24.4
8	42	14	360	1	1	1	23.3	23.3	23.3	23.3	23.3
8	42	14	360	1	1	1	23.0	23.0	23.0	23.0	23.0
SUM:								7.8	12.8	−3.0	39.6
COEFF:								0.49	0.80	−0.19	2.48

$$\widehat{E}_i = \frac{1}{\frac{1}{2}nm}\sum_{j=1}^{nm} h_{ij}y_j$$

$$b_i = \frac{\widehat{E}_i}{2} = \frac{1}{nm}\sum_{j=1}^{nm} h_{ij}y_j$$

The least squares estimates of any two-way interaction coefficients are computed in a similar fashion:

$$b_{ij} = \frac{\widehat{E}_{ij}}{2} = \frac{1}{nm}\sum_{k=1}^{nm} h_{ik}h_{jk}y_{jk}$$

Table 5.4 illustrates the computations of some of the coefficient estimates for the volume yield data.

The only reason these simple formulas for the least squares estimates of the coefficients work is because this is a two-level, balanced, orthogonal design, where each input factor is Helmert-coded.

Figure 5.1 shows the output from the "Fit Model" function of the software package JMP 8.0. Note the agreement between the "Parameter Estimates" section of the Figure and the results in Table 5.4.

Now that the estimates of the first order approximation model are obtained, the next stage is to decide which input variables are likely to have a non-zero effect on the response. In Fig. 5.1, in the Parameter Estimates section, along with the "Term" and "Estimate" columns are three others: Std Error, t Ratio, and Prob >|t|. The Std Error (SE) column is the standard error of the estimated coefficient, or parameter estimate, given by:

$$SE = \sqrt{\frac{MS_{error}}{nm}} = \sqrt{\frac{MS_{error}}{n2^k}}$$

The t Ratio is:

$$t = \frac{b}{SE}$$

where b represents the estimate of the coefficient for the term in question. The formula for SE only has this simple form due to Helmert coding.

Finally, the column labeled "Prob >|t|" is the p-value for testing the hypothesis that the coefficient is actually 0. P-values less than some pre-specified level (usually 0.05) are considered statistically significant, meaning that the actual coefficient does not appear to be 0. Any terms for which the p-value is above the pre-determined threshold are candidates to be excluded from the model.

Figure 5.2 shows Minitab 16 output for the same data and model.

Figure 5.3 shows output (and SAS code) from SAS 9.2

Figure 5.4 shows the output and code for R 3.0.1

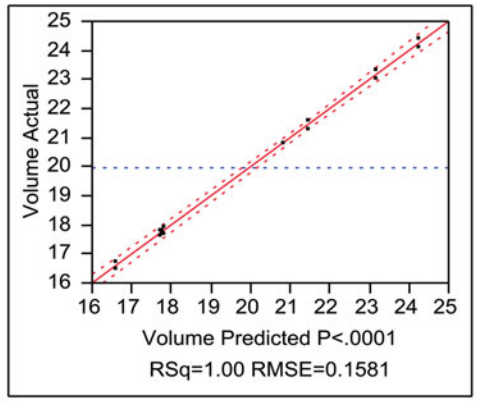
Notice that JMP calls $\sqrt{MS_{error}}$ the "Root Mean Square Error", Minitab calls it "S", SAS calls it "Root MSE", and R calls it "Residual standard error".

In examining the fit of this model, we notice several things. One is that all the three factors, H_1, H_2, and H_3, are significant at the 0.05 level. Thus it does not appear that any of the three are candidates to be ignored. Another is that the interaction terms, H_2*H_3 and H_1*H_2*H_3, are not significant at the 0.05 level. The fact that these interactions are not significant does not imply that any of the factors included in the interaction are not important to control. Yet another thing is that the adjusted R^2 is relatively high (0.99674), indicating that possibly the first order polynomial may be a good enough approximation.

We notice that the run with the highest average volume is run 7, with $H_1 = +1$ (Temperature $= 42^oC$) $H_2 = +1$ (Pressure $= 14$ kP) and $H_3 = -1$ (Time $= 240$ s). The average volume was 24.25 L, which is an improvement over the current volume of 20.0 L. Notice also that the predicted value of volume for run 7 conditions is:

$$\begin{aligned}\text{Volume} = {} & 19.9 + 0.487(+1) + 0.800(+1) - 0.188(-1) + 2.48(+1) - 0.363(-1) - 0.0750(-1) \\ & + 0.0750(-1) = \mathbf{24.25}\end{aligned}$$

Response Volume
Actual by Predicted Plot



Summary of Fit

RSquare	0.998262
RSquare Adj	0.996742
Root Mean Square Error	0.158114
Mean of Response	19.9375
Observations (or Sum Wgts)	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	7	114.89750	16.4139	656.5571
Error	8	0.20000	0.0250	**Prob > F**
C. Total	15	115.09750		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	19.9375	0.039528	504.38	<.0001*
H1	0.4875	0.039528	12.33	<.0001*
H2	0.8	0.039528	20.24	<.0001*
H3	-0.1875	0.039528	-4.74	0.0015*
H1*H2	2.475	0.039528	62.61	<.0001*
H1*H3	-0.3625	0.039528	-9.17	<.0001*
H2*H3	-0.075	0.039528	-1.90	0.0943
H1*H2*H3	0.075	0.039528	1.90	0.0943

Fig. 5.1 JMP output—volume yield first order model

```
Welcome to Minitab, press F1 for help.
```

Regression Analysis: Volume versus H1, H2, H3, H12, H13, H23, H123

```
The regression equation is
Volume = 19.9 + 0.487 H1 + 0.800 H2 - 0.188 H3 + 2.48 H12 - 0.363 H13
         - 0.0750 H23 + 0.0750 H123


Predictor      Coef  SE Coef       T      P
Constant    19.9375   0.0395  504.38  0.000
H1          0.48750  0.03953   12.33  0.000
H2          0.80000  0.03953   20.24  0.000
H3         -0.18750  0.03953   -4.74  0.001
H12         2.47500  0.03953   62.61  0.000
H13        -0.36250  0.03953   -9.17  0.000
H23        -0.07500  0.03953   -1.90  0.094
H123        0.07500  0.03953    1.90  0.094


S = 0.158114   R-Sq = 99.8%   R-Sq(adj) = 99.7%


Analysis of Variance

Source          DF       SS      MS       F      P
Regression       7  114.898  16.414  656.56  0.000
Residual Error   8    0.200   0.025
Total           15  115.098


Source  DF  Seq SS
H1       1   3.802
H2       1  10.240
H3       1   0.563
H12      1  98.010
H13      1   2.103
H23      1   0.090
H123     1   0.090
```

Fig. 5.2 Minitab output—volume yield first order model

It turns out that the least squares-based predicted value of the response (also called the least-squares mean) for any run is actually the arithmetic average response for that run, provided the experiment is balanced. This is another convenient consequence of the two-level experimental design. Many software systems provide least squares means as outputs.

Since this EAS had no specific goal for volume (other than to make it as high as possible), it may be that the only additional experimentation he would want to perform is confirmatory, namely, replicates of run 7, $H_1 = +1$ (Temp $= 42^oC$), $H_2 = +1$ (Pressure $= 14$ kP) and $H_3 = -1$ (Time $= 240$ s). Upon completing another two replicates, the EAS obtained the values 24.4 L and 24.5 L. These two values, and their average, 24.45 L, were close enough to the prediction that the EAS decided that the best conditions were at Temperature $= 42^oC$, Pressure $= 14$ kP, and Time $= 240$ s.

It is quite possible that the chosen operating conditions are suboptimal. The presumption is that the chosen range of input factors in which the experiment was conducted at least contains the optimal point. Secondly, there is no way using the data gathered in this two-level experiment to decide that any higher order terms would improve the polynomial approximation. It might be helpful to perform the experiment at an intermediate point; something between the "corners" as defined by the high and low levels of the input factors. Recall that the factors in natural units were Helmert-coded by

subtracting the midpoint values from the high and low values of each factor. The midpoints for each factor would then be coded (or mapped) to the value 0. The "center" of the experimental space in coded units would then be at $H_i = 0, i = 1, 2, 3$. If the average response at this center point is "close' to the predicted value from the first order model, then it is more believable that no higher order terms are necessary. Note that at the Helmert-coded center point, the predicted value is the intercept estimate, which was ~19.94 L.

The center point conditions for the volume yield experiment are Temperature = 37°C, Pressure = 12 kP, Time = 300 s. The EAS obtained two replicate values for the center point: 20.3 L and 19.6 L, for an average response of 19.95 L. This is very close to the intercept estimate of 19.94 L. Therefore, evidence indicates that the first order model is adequate within the range of the input factors in which the experiment was conducted.

```
libname stuff 'H:\Experimentation for Design & Validation\Data &
Analyses';

data calc;

  set stuff.d20130812_example_5_1_volume;
/*************************************************************
*                                                            *
* variables: Run, Temp, Pressure, Time, H1, H2, H3, Volume   *
*                                                            *
*************************************************************/

  run;

proc glm data=calc;
  model Volume = H1 H2 H3 H1*H2 H1*H3 H2*H3 H1*H2*H3/est;
  run;

The SAS System          08:19 Monday, August 12, 2013   8

                              The GLM Procedure

                    General Form of Estimable Functions

                        Effect        Coefficients

                        Intercept     L1

                        H1            L2

                        H2            L3

                        H3            L4

                        H1*H2         L5

                        H1*H3         L6

                        H2*H3         L7

                        H1*H2*H3      L8
```

Fig. 5.3 SAS output—volume yield first order model

```
                                The SAS System          08:19 Monday, August 12, 2013   9

                                The GLM Procedure

Dependent Variable: Volume   Volume

                                                Sum of
Source                          DF             Squares      Mean Square    F Value    Pr > F

Model                            7          114.8975000       16.4139286     656.56    <.0001

Error                            8            0.2000000        0.0250000

Corrected Total                 15          115.0975000

                  R-Square     Coeff Var      Root MSE    Volume Mean

                  0.998262      0.793048      0.158114       19.93750

Source                          DF       Type I SS     Mean Square    F Value    Pr > F

H1                               1       3.80250000      3.80250000     152.10    <.0001
H2                               1      10.24000000     10.24000000     409.60    <.0001
H3                               1       0.56250000      0.56250000      22.50    0.0015
H1*H2                            1      98.01000000     98.01000000    3920.40    <.0001
H1*H3                            1       2.10250000      2.10250000      84.10    <.0001
H2*H3                            1       0.09000000      0.09000000       3.60    0.0943
H1*H2*H3                         1       0.09000000      0.09000000       3.60    0.0943

Source                          DF      Type III SS     Mean Square    F Value    Pr > F

H1                               1       3.80250000      3.80250000     152.10    <.0001
H2                               1      10.24000000     10.24000000     409.60    <.0001
H3                               1       0.56250000      0.56250000      22.50    0.0015
H1*H2                            1      98.01000000     98.01000000    3920.40    <.0001
H1*H3                            1       2.10250000      2.10250000      84.10    <.0001
H2*H3                            1       0.09000000      0.09000000       3.60    0.0943
H1*H2*H3                         1       0.09000000      0.09000000       3.60    0.0943

                                                Standard
Parameter                    Estimate             Error     t Value    Pr > |t|

Intercept                19.93750000       0.03952847      504.38      <.0001
H1                        0.48750000       0.03952847       12.33      <.0001
H2                        0.80000000       0.03952847       20.24      <.0001
H3                       -0.18750000       0.03952847       -4.74      0.0015
H1*H2                     2.47500000       0.03952847       62.61      <.0001
H1*H3                    -0.36250000       0.03952847       -9.17      <.0001
H2*H3                    -0.07500000       0.03952847       -1.90      0.0943
H1*H2*H3                  0.07500000       0.03952847        1.90      0.0943
```

Fig. 5.3 (continued)

Non-continuously Valued Input Factors and Multiple Comparisons

Often factors are ordinal or even nominal valued. That is, a setting on a machine might have only discrete, albeit ordered, values (slow, medium, and fast speeds, for example) or a factor's levels may have no order at all (colors, such as white or red, for example). In these cases, Helmert coding can still be used, but the interpretation of the output is very different. Instead of fitting the model to be predictive, the model is discriminatory. That is, rather than interest in estimating the model coefficients, we would be more interested in estimating the effect, and testing whether the effect for a given term is significantly different from zero. Most the models we have created so far were predictive polynomial approximations to the relationship between the response and the factors. The number of parameters associated with a factor or an interaction was one, and it was interpreted as a

```
> setwd("H:\\Personal Data\\Experimentation for Design & Validation\\Data &
Analyses\\")
> df1 <- read.csv("20130812_example_5_1_volume.csv")
>
> attach(df1)
>
> first_order <- lm(Volume ~ H1 + H2 + H3 + H1*H2 + H2*H3 + H2*H3 + H1*H2*H3)
>
> summary(first_order)

Call:
lm(formula = Volume ~ H1 + H2 + H3 + H1 * H2 + H2 * H3 + H2 *
    H3 + H1 * H2 * H3)

Residuals:
   Min     1Q Median     3Q    Max
 -0.15  -0.10   0.00   0.10   0.15

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 19.93750    0.03953 504.383  < 2e-16 ***
H1           0.48750    0.03953  12.333 1.74e-06 ***
H2           0.80000    0.03953  20.239 3.71e-08 ***
H3          -0.18750    0.03953  -4.743  0.00146 **
H1:H2        2.47500    0.03953  62.613  4.71e-12 ***
H2:H3       -0.07500    0.03953  -1.897  0.09435 .
H1:H3       -0.36250    0.03953  -9.171   1.61e-05 ***
H1:H2:H3  0.07500    0.03953   1.897   0.09435 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1581 on 8 degrees of freedom
Multiple R-squared:  0.9983,    Adjusted R-squared:  0.9967
F-statistic: 656.6 on 7 and 8 DF,  p-value: 2.132e-10

>
> detach(df1)
```

Fig. 5.4 R output—volume yield first order model

slope. When factors are discrete, the notion of slope does not really apply. Even in the case where a discrete factor has only two levels, the effect, as described in the beginning of this chapter, is a more meaningful measure than slope. In Figs. 5.2 and 5.3, the reader may have noticed tables with columns labeled "source" and "DF". The "source was the factor or regressor variable, and DF was the degrees of freedom associated with the factor, and it was always one. In the case of discrete factors, the degrees of freedom associated with the factor are the number of levels −1. The degrees of freedom for a factor are actually the number or parameters required to represent the factor's effect. In the case of

two-level factors, and in fact for any continuously valued factor, only one parameter is required. If a factor is continuous, then the parameter can be interpreted as a slope. In the case of discrete factors, the parameters associated with those factors are only useful inasmuch as they are used to assess the factor's effect.

Usually, in the case of discrete factors, the EAS wants to know more specifically which combinations of levels for a factor differ significantly from each other. Tests of significance comparing specific combinations of factor levels are called multiple comparison tests (Montgomery 2001). These tests are generally only applied if the overall effect is significant, and are most useful when a factor has more than two levels. The tests are designed to control for inflating the chance of concluding that a significant difference between pairs of specific combinations of factor levels exists, when in fact it does not. One such test is the Tukey-Kramer test (Montgomery, ref.cit.), or sometimes referred to as Tukey's Honestly Significant Difference, or HSD, test (Adler 2010). Figure 5.5 shows R code for an example with one discrete factor having three levels (called "first", "second", and "third"). The R Output for the example, including the Tukey HSD test, is shown in Fig. 5.6. The columns labeled "lwr" and "upr" in the output for the TukeyHSD function are the lower and upper limits of simultaneous 95 % confidence intervals for the differences between the average response at the particular levels being compared. By "simultaneous" we mean that jointly the confidence level for all the intervals is 95 %. The p adj column is the p-value, adjusted to account for multiple comparisons.

In this example, the factor fF1 is significant ($p = 1.111e\text{-}05$). While level "second" differs significantly from level "third" ($p = 0.0000122$) and from level "first" ($p = 0.0007988$), level "first" does not differ significantly from level "third" ($p = 0.3909875$).

If all the factors in the experiment, continuous or non-continuous, are restricted to two levels, then we can take advantage of all the computational efficiencies afforded by 2^k factorial designs. So, the designs and their associated characteristics are not restricted to only continuously valued factors. In fact, both discrete and continuous factors may be included in a single experimental design.

It is possible that in addition to controlling levels of discrete and continuous factors, the EAS may find that there are one or more continuously-valued quantities that vary without control, but which can

Fig. 5.5 Example R code with Tukey HSD function

```
setwd("H:\\Personal Data\\Experimentation for Design & Validation\\Data &
Analyses\\")
df1 <- read.csv("20150121 Example Multiple Comparisons.csv")

attach(df1)

fF1 <- factor(F1, 1:3)
levels(fF1) <- c("first","second","third")

model123 <- aov(Y ~ fF1, na.action=na.omit)
#
# TukeyHSD requires an aov object as its first argument
#
TukeyHSD(x=model123,which="fF1",ordered=TRUE)

summary(model123)
#
# Note that the anova function will be the same as summary(model123)
# if model123 is an aov object
#

anova(model123)
detach(df1)
```

```
> setwd("H:\\Personal Data\\Experimentation for Design & Validation\\Data &
Analyses\\")
> df1 <- read.csv("20150121 Example Multiple Comparisons.csv")
>
> attach(df1)
>
> fF1 <- factor(F1, 1:3)
> levels(fF1) <- c("first","second","third")
>
>
> model123 <- aov(Y ~ fF1, na.action=na.omit)
> #
> # TukeyHSD requires an aov object as its first argument
> #
> TukeyHSD(x=model123,which="fF1",ordered=TRUE)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Y ~ fF1, na.action = na.omit)

$fF1
                diff        lwr        upr     p adj
first-third        1 -0.8374262   2.837426 0.3909875
second-third       4  2.1625738   5.837426 0.0000122
second-first       3  1.1625738   4.837426 0.0007988
>
> summary(model123)
             Df  Sum Sq Mean Sq  F value   Pr(>F)
fF1           2   130.0   65.00    15.15 1.11e-05 ***
Residuals    42   180.2    4.29
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #
> # Note that the anova function will be the same as summary(model123)
> # if model123 is an aov object
> #
>
```

Fig. 5.6 R output for TukeyHSD function

```
> anova(model123)
Analysis of Variance Table

Response: Y
           Df  Sum Sq  Mean Sq  F value   Pr(>F)
fF1         2  130.00   65.00   15.152    1.111e-05 ***
Residuals 42  180.18    4.29
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
> detach(df1)
```

In this example, the factor fF1 is significant (p = 1.111e-05). While level "second" differs significantly from level "third" (p = 0.0000122) and from level "first" (p = 0.0007988), level "first" does not differ significantly from level "third" (p = 0.3909875).

Fig. 5.6 (continued)

be measured. For example, the ambient temperature may not be completely controllable, and it may have some effect on experimental response variables. A version of ANOVA, called Analysis of Covariance (ANCOVA) was developed to allow the experimenter to compare the levels of discrete factors while compensating for a free-varying "covariate", which may attenuate the response signal. ANCOVA is a sort of combined ANOVA and regression. It is not the same as including a continuously varying but uncontrolled input in a regression model; rather ANCOVA is a means of adjusting the test of significance for discrete factors, to compensate for unwanted effects of the covariates. Inasmuch as this text focuses mostly on constructing predictive models, and not so much on determining the significance of discrete factors, no more will be said about ANCOVA. The Montgomery text already cited several times has an excellent coverage of this topic.

Matrix Form

Suppose **H** is the matrix of all the columns relating to the intercept and each input factor coefficient, including each cross-product terms (obtained by row-wise multiplication of the factors in the cross-product terms). Furthermore, let **Y** represent the vector of all the Volume (response) results. They would look like Fig. 5.7.

The first order model can be represented by the vector equation:

$$\mathbf{Y} = \mathbf{H}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

Fig. 5.7 Matrix representation of volume yield experiment

			H Matrix					Y vector
Intercept	H1	H2	H3	H1H2	H1H3	H2H3	H1H2H3	y=Volume
1	-1	-1	-1	1	1	1	-1	20.8
1	-1	-1	-1	1	1	1	-1	20.8
1	-1	-1	1	1	-1	-1	1	21.6
1	-1	-1	1	1	-1	-1	1	21.3
1	-1	1	-1	-1	1	-1	1	17.8
1	-1	1	-1	-1	1	-1	1	17.7
1	-1	1	1	-1	-1	1	-1	17.9
1	-1	1	1	-1	-1	1	-1	17.7
1	1	-1	-1	-1	-1	1	1	17.6
1	1	-1	-1	-1	-1	1	1	17.8
1	1	-1	1	-1	1	-1	-1	16.5
1	1	-1	1	-1	1	-1	-1	16.7
1	1	1	-1	1	-1	-1	-1	24.1
1	1	1	-1	1	-1	-1	-1	24.4
1	1	1	1	1	1	1	1	23.3
1	1	1	1	1	1	1	1	23.0

where:

$$\boldsymbol{\theta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \gamma_{12} \\ \gamma_{13} \\ \gamma_{23} \\ \delta_{123} \end{bmatrix}$$

and $\boldsymbol{\varepsilon}$ is a $nm \times 1$ vector of random noise variables. If $\widehat{\boldsymbol{\theta}}$ represents a vector of the ordinary least squares estimates of the model coefficients, then it can be expressed in the vector equation:

$$\widehat{\boldsymbol{\theta}} = \left[\boldsymbol{H}'\boldsymbol{H}\right]^{-1}\boldsymbol{H}'\boldsymbol{Y}$$

That is, $\widehat{\boldsymbol{\theta}}$ is the solution to the least squares minimization problem (Draper and Smith 1998), namely to find the values of the coefficients that minimize the sum of squared errors (*SSE*).

Recall the formula for the standard error of predicted values:

$$SE\left(\widehat{y}|\boldsymbol{h}_0\right) = s\sqrt{\boldsymbol{h}_0'\left[\boldsymbol{H}'\boldsymbol{H}\right]^{-1}\boldsymbol{h}_0}$$

where s is the root mean square error, and $\boldsymbol{h_0}$ represents a particular design point (i.e., a value corresponding to each column of the **H** matrix) in the Helmert-coded input variable space. Once again the Helmert coding simplifies the computations. The matrix $\boldsymbol{H}'\boldsymbol{H}$ has diagonal elements all equal to np, where p = number of parameters in the model = number of columns in the matrix $\boldsymbol{H}$, and all off-diagonal elements equal to 0. The inverse of such a matrix is simply the reciprocal of the diagonal elements on its diagonal and zero everywhere else. That means:

$$SE\left(\widehat{y}\middle|\boldsymbol{h}_0\right) = s\sqrt{\boldsymbol{h}_0'\left[\boldsymbol{H}'\boldsymbol{H}\right]^{-1}\boldsymbol{h}_0} = \frac{s}{\sqrt{n}}$$

For run 7, $\boldsymbol{h}_0' = [+1, +1, +1, -1, +1, -1, -1, -1]$, and

$$SE\left(\widehat{y}\middle|\boldsymbol{h}_0\right) = \frac{s}{\sqrt{2}}$$

With $s \approx 0.1581$, the 99 % lower confidence bound on the predicted value for run 7 is:

$$\widehat{y} - t(0.99, 16-8)SE\left(\widehat{y}\middle|h_0\right) \approx 24.25 - 2.896*\frac{0.1581}{\sqrt{2}} \approx 23.93$$

So, now we have seen three different ways to compute the least squares estimates of the first order polynomial coefficients using data from a balanced, 2^k factorial experiment:

1. Compute the "effect" in terms of differences in average response at the high and low levels of the input factor, and then dividing by 2;
2. Using the coded column for the input factor, multiply it row-wise with the corresponding response variable values, and then sum the products; divide by the total number of response values;
3. Use the matrix/vector solution to the least squares minimization problem;

All three of these methods yield the same answer. Method 2 only works for Helmert-coded, balanced, orthogonal two-level experiments.

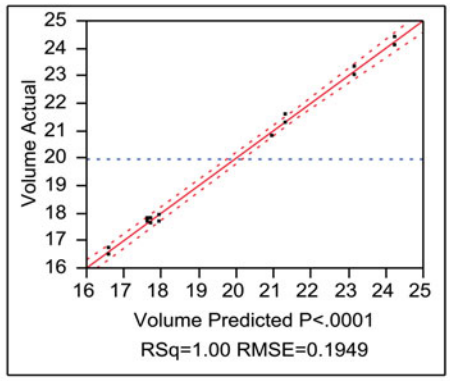
Reducing the Model

Inasmuch as two interaction terms in the Volume Yield model were not significant (i.e., the p-values for the t-ratios of their coefficient estimates were greater than 0.05), we might consider what would happen if we "reduced" the model, excluding the H_2H_3 and $H_1H_2H_3$ terms, which both had associated p-values of approximately 0.09435 (Fig. 5.4). The JMP output for the reduced model is shown in Fig. 5.8.

Apart from missing the H_2H_3 and $H_1H_2H_3$ terms, the coefficients in the new model are identical to those of the corresponding terms in the full model. The reduced model has a lower adjusted R^2 (0.9950 instead of 0.9967), a higher root means square error (0.1949 instead of 0.1581), and therefore correspondingly higher standard errors for the coefficients (0.0487 instead of 0.0395). The only question is whether the reduced model gives better predictions than the full model. For run 7, the predicted value of volume with the full model was 24.25 L. With the reduced model, it is also about 24.25 L. The two models agree very closely at all the runs. Table 5.5 shows the predicted values at each of the 8 runs. Of course, since the intercept estimates are identical for the two models, the predicted values at the center point run, $H_1 = H_2 = H_3 = 0$, are the same.

The only way to assess which model yields superior predictions is to replicate runs at which the two predictions (full and reduced model) do not agree. In particular, since the H_2H_3 and $H_1H_2H_3$ interactions were dropped in the reduced model, it might be instructional to replicate a run in which all three coded variables had the same sign, and a run in which one variable differed in sign from the other two. Run 1 ($H_1 = H_2 = H_3 = -1$) and run 8 ($H_1 = H_2 = H_3 = +1$) are the two runs where all variables have the same sign. Any of the other runs would satisfy the 2 out of 3 runs having different signs. However, note that the predicted values for runs 5, 6, and 7 are identical, so the information

Response Volume
Actual by Predicted Plot



Summary of Fit

RSquare	0.996698
RSquare Adj	0.995048
Root Mean Square Error	0.194936
Mean of Response	19.9375
Observations (or Sum Wgts)	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	114.71750	22.9435	603.7763
Error	10	0.38000	0.0380	Prob > F
C. Total	15	115.09750		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	19.9375	0.048734	409.11	<.0001*
H1	0.4875	0.048734	10.00	<.0001*
H2	0.8	0.048734	16.42	<.0001*
H3	-0.1875	0.048734	-3.85	0.0032*
H1*H2	2.475	0.048734	50.79	<.0001*
H1*H3	-0.3625	0.048734	-7.44	<.0001*

Fig. 5.8 JMP output—reduced model—volume yield first order model

Table 5.5 Predicted values, full and reduced models

Run	Full model	Reduced model
1	20.80	20.95
2	21.45	21.30
3	17.75	17.60
4	17.80	17.95
5	17.70	17.70
6	16.60	16.60
7	24.25	24.25
8	23.15	23.15

about which model yields a better prediction would not be available. So, in order to assess which model yields better predictions, runs 2, 3, or 4 are the best candidates.

Before obtaining any further experimental data, however, examine the differences in the predictions. Are those differences materially different? Furthermore, does it matter, since the maximum prediction for both models is 24.25 L at "run 7" conditions? Since the function being maximized is linear in the parameters, and the only constraints on the inputs are that they fall within specified ranges (coded to the range −1 to +1), there are no intermediate settings of the input variables that will increase the polynomial approximation to the yield function. Therefore, given that the center point data did not indicate any higher order effects beyond the linear model, and given that the reduced model did not yield any alteration in what appeared to be the optimal conditions, the EAS may conclude that the process should be run at "run 7" conditions, namely temperature = 42°C, pressure = 14 kP, and time = 240 s. The 99 % lower bound on expected volume yield is:

$$\widehat{y} - t(0.99, 16-6)SE\left(\widehat{y}\middle|h_0\right) \approx 24.25 - 2.764*\frac{0.1949}{\sqrt{2}} \approx 23.92$$

based on the reduced model.

To reiterate the modeling, analysis, and decision process:

1. Choose an objective, such as maximizing yield, minimizing waste, or obtaining a measurement that exceeds some lower limit. In all the cases discussed in this chapter, there is a single output, or response variable, and that is can be expressed as a real number.
2. Choose a set of candidate input variables, which are those controllable factors that are suspected to have some effect on the output response.
3. Choose operating ranges for each of the controllable input factors.
4. Fit the full, first order polynomial to Helmert-coded input variables.
5. Obtain some runs at the center point.
6. Compare the average of the center point runs to the intercept. Are they "close"?
7. Determine which, if any, terms in the full model have coefficients that are not significantly different from zero.
8. Compute predicted values for a reduced model, which only contains terms with significant coefficients.
9. Find the run conditions that provide the "best" predicted output response. Compute bounds for the prediction (lower, upper, or both).
10. If possible, consider obtaining some additional data using conditions that were not included in the original experiment (other than the center point). Compute a predicted value from the model, and compare it to the actual average response from the additional run(s). Are they "close"? Does the prediction "seem" reasonable?
11. If the predicted values seem close to the actual results obtained from additional runs, then use the run conditions that yield the "best" predicted value.
12. If the average value of the center point runs is not close to the predicted value of the response, or the actual values of additional runs are not close to predicted values, then a higher order approximation may be necessary.

Keep in mind that there are potentially other phenomena that may affect the modeling and analysis process. There can be budgetary as well as physical constraints that limit your ability to experiment. It is possible that some input factors may only have either an operating minimum or maximum, so that choosing a range may be challenging. Of course, measurement systems for both output and input variables may constrain the ability to know the true optimal value or choose optimal operating/design conditions. And noise levels may make confidence limits wider than desirable for a given sample size.

Key Points

- System features can be thought of as input factors that affect the value of a performance measure, called the response variable.
- Factorial experiments are sets of factor values used to assess the effects of each factor on the average value of the response variable.
- Two-level factorial experiments are used to find first-order approximating polynomials. They can tell you if increasing a given factor increases, decreases, or does not change the average value of the response.
- Interaction effects are assessed by including two-way cross-products of input factors into the model.

Exercises and Questions

1. What order polynomial can be fit using data from a two-level factorial experiment?
2. How can a center point run be useful?
3. What analyses can be used for discrete factors? How do those analyses differ from continuous factors?

Chapter 6
Fractional Factorial Designs

In the two-level experiment with k input factors, there are 2^k possible runs. For $k = 3$, there are only eight runs. But even with five factors, $2^5 = 32$ runs may be prohibitively expensive and time consuming. The question is whether it is possible to use some subset of all possible runs and still be able to fit a first order polynomial approximation to the response function. The answer is yes, but you must give up some resolution in your approximation. The amount of resolution lost is generally in terms of which interaction effects which can be included in your model.

The process of selecting the subset of runs is called fractional replication, and the experimental designs resulting from the process are called fractional factorials. The term comes from the fact that selected fractions of the total number of runs are of the form $\frac{1}{2^p}$, where $p = 1, 2, 3, \ldots$ So, for example, if there are five factors over which the experiment is to be performed, then there are $2^5 = 32$ possible runs, and a $\frac{1}{2^1}$ or one-half fraction would require only 16 of those runs. This section will deal with a sort of taxonomy for fractional factorial designs, in terms of the amount of information lost, together with a method for choosing the particular fraction of the total number of possible runs. The method of choosing those runs will preserve balance and orthogonality, the two essential characteristics that allow for the unambiguous estimation of model coefficients.

Resolution

The degree to which information is lost when a fraction of the total number of runs are used in an experiment is called resolution. A complete, full factorial design has full resolution. The next resolution level most commonly employed is called Resolution V (Roman numeral 5), or ResV. The resolution in a design limits the order of interaction, or cross-product terms, which can be included in a model. For ResV designs, the approximating first order polynomial can have all possible two-way cross-product, or interaction, terms. All three-way or higher order cross-product terms are what is called "aliased" with two-way terms. That means the coefficient estimates for any three-way terms will be identical to the estimate for some two-way term. The other resolution levels we will discuss are ResIV (where single factor terms are not aliased with each other, although they are aliased with higher order cross-product terms, and two-way cross-products are aliased with other two-way cross-products and higher order cross-products, but no cross-product terms are aliased with any single factor terms) and ResIII (where two-way and higher order cross-product terms are all aliased with single factor terms, but single-factor terms are not aliased with each other). Generally, a design is of resolution K if the lowest order term that any k-order term is aliased with is K-k. So for example, in a ResV (K = 5) design, the lowest order term

S.A. Pardo, *Empirical Modeling and Data Analysis for Engineers and Applied Scientists*,
DOI 10.1007/978-3-319-32768-6_6

Table 6.1 An experiment with aliasing

Run	H1	H2	H3	Y
1	−1	−1	−1	y_1
2	−1	1	−1	y_2
3	1	−1	1	y_3
4	1	1	1	y_4

any first-order term will be aliased with is a $5 - 1 =$ fourth order cross-product term, second order terms will be aliased with $5 - 2 =$ third-order terms, and of course third-order terms will be aliased with $5 - 3 =$ second-order terms.

First, we will illustrate the idea of aliasing. Secondly we will describe a method for dividing the 2^k runs into 2^p subsets, or fractional designs, in such a way as to preserve balance and orthogonality. Then we will discuss a method for determining p in order to have a particular resolution.

Aliasing

Suppose we had three Helmert-coded input factors, call them H_1, H_2, and H_3. Suppose that we had chosen to make only four runs, as shown in Table 6.1. If we did this experiment, then the estimate for the coefficients of factors H_1 and H_3 would be identical. They are said to be aliased. Suppose we tried to fit the model:

$$y = \beta_0 + \beta_1 H_1 + \beta_2 H_2 + \beta_3 H_3 + \varepsilon$$

Using the information in Table 6.1, we would have the following estimates of the coefficients β_1 and β_3:

$$b_1 = \frac{(-1)y_1 + (-1)y_2 + (+1)y_3 + (+1)y_4}{4}$$

and

$$b_3 = \frac{(-1)y_1 + (-1)y_2 + (+1)y_3 + (+1)y_4}{4}$$

So the estimates would be identical.

If we tried to use a computer program to compute the estimates of the model coefficients, we would get an error message, telling us about a singular matrix or collinearity or some other message that indicates the least squares estimates for the coefficients could not be computed. So, in models fit to data from fractional factorial experiments, care must be taken to only include terms in the model that are not aliased with each other.

In all fractional factorials, the single-factor terms of the models will be aliased with some other terms. We would like to exclude from the model the higher-order cross-product terms that are aliased with lower-order terms, so that the model would give us the best approximation possible. Generally, the higher the order of the cross-product, the less important it is in terms of approximation. The largest fraction of a 2^k design that would be able to preserve balance and orthogonality would be 2^{-1} or one half of the total number of runs. To select such a fraction, write down the runs of a complete or full 2^{k-1} design, using Helmert-coded levels. Then compute the levels for the kth factor by multiplying across the columns of the $k - 1$ factors. For example, suppose we wanted a one-half fraction of a 2^3 experiment. The rows of the $2^{(3-1)} = 2^2$ full factorial would look like Table 6.2:

Table 6.2 Full 2^2 factorial experiment

Run	H1	H2
1	−1	−1
2	−1	1
3	1	−1
4	1	1

Table 6.3 One-half fraction of the 2^3 design

Run	H1	H2	H3 = H1H2
1	−1	−1	1
2	−1	1	−1
3	1	−1	−1
4	1	1	1

To add the third factor, equate it to the interaction (cross-product) term W_1W_2:

Table 6.3 illustrates this fractional design:

Thus, in this fractional design, the third factor, H_3, is aliased with the H_1H_2 cross-product. Similarly, you can verify that the single-factor term W1 is aliased with the W_2W_3 cross-product, and H_2 is aliased with H_1H_3. Thus, the only model that can be fit using this experiment that would include all three input variables is:

$$y = \beta_0 + \beta_1 H_1 + \beta_2 H_2 + \beta_3 H_3 + \varepsilon$$

It would be possible to fit instead the model:

$$y = \beta_0 + \beta_1 H_1 + \beta_2 H_2 + \beta_{12} H_1 H_2 + \varepsilon$$

In fact, the estimates of β_3 and β_{12} would be identical. Of course, we are hoping that the single-factor term H_3 is actually dominating the two-way cross-product H_1H_2. The tacit assumption is that the higher order the cross-product, the less influence it has on the response variable. Of course, it is possible that this assumption is incorrect.

Generating a Fractional Factorial

To create a 2^{-p} fraction of a 2^k design, first write down in Helmert-coded for all the runs for any subset of $k - p$ factors. Then create a new column for the next factor, by multiplying the coded values for $k - p$ of the previously selected factors. Do this for each additional column, choosing different combinations of the other factor columns.

As an example, consider the 2^{-1} or one-half fraction of a 2^5 design. In a full two-level, five-factor experiment, there would be $2^5 = 32$ runs. In an half-fraction, there would be $\frac{2^5}{2^1} = 16$ runs.

Suppose we label the factors H1, H2, H3, H4, and H5. First, write down the coded columns of a full 2^4 design in, say, H1 through H4. There would be $2^4 = 16$ runs for this design. Then, compute the column for H5 = H1*H2*H3*H4. Table 6.4 illustrates the design.

In this experiment, for example, run 1 would have factors H1–H4 all set to their "low" (−1) level, and H5 set to its "high" (+1) level. The coefficient estimate for H5 would be identical to that of the fourth order cross-product term H1H2H3H4. Similarly, the levels of H1 would be identical to those of the fourth-order cross-product H2H3H4H5. Again, invoking Ockham's razor, we will opt to include the first-order term H1 in our model, and exclude the fourth-order terms. Note that the lowest order

Table 6.4 Half-fraction of a 2^5 design

Run	H1	H2	H3	H4	H5 = H1H2H3H4
1	−1	−1	−1	−1	1
2	−1	−1	−1	1	−1
3	−1	−1	1	−1	−1
4	−1	−1	1	1	1
5	−1	1	−1	−1	−1
6	−1	1	−1	1	1
7	−1	1	1	−1	1
8	−1	1	1	1	−1
9	1	−1	−1	−1	−1
10	1	−1	−1	1	1
11	1	−1	1	−1	1
12	1	−1	1	1	−1
13	1	1	−1	−1	1
14	1	1	−1	1	−1
15	1	1	1	−1	−1
16	1	1	1	1	1

cross-product terms for which any first-order term is aliased is a $5 - 1 =$ fourth order term. Also, the second-order terms are aliased with third-order terms (check, for example, the column for the H1H2 term and for the H3H4H5 column). Thus, the $k =$ second order terms are aliased with the $5 - 2 =$ third-order terms. Since second order cross-products are not aliased with each other or with first order terms, this half-fraction of the 2^5 design is a ResV design.

There is a simply procedure to determine exactly which terms are aliased. First, we chose the H5 term to be aliased with the fourth-order term H1H2H3H4. We obtained the H5 column by multiplying across the rows the columns of H1, H2, H3, and H4. Any column multiplied by itself, row by row, would yield a column of all +1's. Multiplying any column by a column of all +1's would yield the original column. In other words, the column of all +1's acts like a multiplicative identity. Since we chose to alias H5 with H1H2H3H4, we will call the product I = H1H2H3H4H5 (= H5*H5) the *generator* of the design. To find out which other terms any term is aliased with, simply multiply the symbol for that term with the identity, and recognize that any term multiplied by itself will yield a column of all +1's, so that it can simply be ignored. Here are the aliases for our half-fraction design:

$H1*I = H1^2H2H3H4H5 = H2H3H4H5$
$H2*I = H1H2^2H3H4H5 = H1H3H4H5$
$H3*I = H1H2H3^2H4H5 = H1H2H4H5$
$H4*I = H1H2H3H4^2H5 = H1H2H3H5$
$H5*I = H1H2H3H4H5^2 = H1H2H3H4$
$H1H2*I = H1^2H2^2H3H4H5 = H3H4H5$
$H1H3*I = H1^2H2H3^2H4H5 = H2H4H5$
$H1H4*I = H1^2H2H3H4^2H5 = H2H3H5$
$H1H5*I = H1^2H2H3H4H5^2 = H2H3H4$
$H2H3*I = H1H2^2H3^2H4H5 = H1H4H5$
$H2H4*I = H1H2^2H3H4^2H5 = H1H3H5$
$H2H5*I = H1H2^2H3H4H5^2 = H1H3H4$
$H3H4*I = H1H2H3^2H4^2H5 = H1H2H5$
$H3H5*I = H1H2H3^2H4H5^2 = H1H2H4$
$H4H5*I = H1H2H3H4^2H5^2 = H1H2H3$

Table 6.5 Two possible half-fractions of a 2^5 design

Run	H1	H2	H3	H4	H5 = H1H2H3H4	H5(alt) = −H1H2H3H4
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
13	1	1	−1	−1	1	−1
14	1	1	−1	1	−1	1
15	1	1	1	−1	−1	1
16	1	1	1	1	1	−1

Since the term H1H2H3H4H5 = I (column of all +1's) it is actually aliased with the intercept term. Recall from Chap. 4 that the "design" (**H**) matrix had a column of all 1's appended to the columns of the input variables, and that this column was used in the estimation of the intercept.

Now it is easy to see that all the first-order terms are aliased with fourth order terms, and all the second-order terms are aliased with third-order terms, and the fifth-order term is aliased with the intercept. Mr. Ockham would tell us to (at least in the beginning) include the lowest order terms not aliased with each other in our model. Our model would then be:

$$y = \beta_0 + \sum_{i=1}^{r} \beta_i H_i + \sum_{i \neq j} \sum \gamma_{ij} H_i H_j + \varepsilon$$

There are always two one-half fraction designs for any full 2^k design corresponding to ±I. So, we used I = H1H2H3H4H5 to generate the half fraction; we could also have used I = −H1H2H3H4H5. That is, instead of assigning to the H5 column the row-wise product of the columns H1, H2, H3, and H4, we could just as easily taken those products and multiplied them by −1. Table 6.5 shows the original half-fraction with H5 = H1H2H3H4, and an alternate fraction where H5 (alt) = −H1H2H3H4.

Either of these fractions (use either H5 or H5(alt)) would allow us to fit the same model containing the first and second-order cross-product terms. Neither is superior in any statistical sense. It is possible that one of the fractions may be easier to execute for some physical reason. Both fractions are ResV.

Aside from controlling the nature of aliasing, the process we have used to select a half-fraction preserves balance and orthogonality. Thus the coefficient estimation procedure is identical for fractional and full two-level designs. The only thing that has changed is the model which we can fit to the data. Clearly, using fractional designs induces a loss in information; we are hoping that Ockham was right, and that we are losing the least important bits of information as a tradeoff for economy.

Table 6.6 A one-quarter fraction

Run	H1	H2	H3	H4	H5 = H1H2H3	H6 = H2H3H4
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
13	1	1	−1	−1	−1	1
14	1	1	−1	1	−1	−1
15	1	1	1	−1	1	−1
16	1	1	1	1	1	1

Generating a One-Quarter Fraction

Suppose we had not five factors, but six factors with which to experiment. A full factorial in six factors would require $2^6 = 64$ runs. A one-half fraction would require $2^{6-1} = 32$ runs. If we could only afford 16 runs, we would need to generate a 2^{-2} or one-quarter fraction of the total 64 runs. The method is very similar to that for generating the 2^{-1} fraction. First, write down the runs for a full 2^4 design in factors H1, H2, H3, and H4 (the full factorial with the required number of runs). Then assign to another column for H5 the levels (in Helmert-coded form) computed from the row-wise product of H1, H2, and H3. Then assign to yet another column for H6 the product of H2, H3, and H4. In the quarter-fraction design, there are $p = 2$ *generators*. The result will be a 16-run design with six factors. Table 6.6 shows the design.

To determine the aliases of each other term, multiply the term by each *generator*, H1H2H3H5 and H2H3H4H6. Furthermore, multiply each term by the product of H1H2H3H5 and H2H3H4H6 = H1H4H5H6. Just as in the case of the half-fraction, we define identity products based on the terms we chose to alias, namely I = H1H2H3H5 = H2H3H4H6. However, since I*I = I, we must also include in the definition of the identity the product of H1H2H3H5 and H2H3H4H6, or H1H2H3H5*H2H3H4H6 = $H1H2^2H3^2H4H5H6$ = H1H4H5H6. So, the full definition of the identity cross-product is:

I = H1H2H3H5 = H2H3H4H6 = H1H4H5H6.

To determine the aliases of each term, multiply those terms by the identities:

H1*I = H2H3H5 = H1H2H3H4H6 = H4H5H6
H2*I = H1H3H5 = H3H4H6 = H1H2H4H5H6
H3*I = H1H2H5 = H2H4H6 = H1H2H3H5H6
H4*I = H1H2H3H4H5 = H2H3H6 = H1H5H6
H5*I = H1H2H3 = H2H3H4H5H6 = H1H4H6
H6*I = H1H2H3H5H6 = H2H3H4 = H1H4H5

The second-order products have an interesting alias structure. For example:

H1H2*I = H3H5 = H1H3H4H6 = H2H4H5H6

Notice that this second-order product term is aliased with another second-order product. That means, in our model, we cannot have both H1H2 and H3H5.

The design of this one-quarter fraction is ResIV, since for the k = second-order product terms, they are aliased with another K − k = 4 − 2 = 2, or second-order cross-product. The problem is that now the modeler must chose only a subset of the second-order cross-product terms to include in the model. Sometimes, physical laws and known phenomena may dictate which cross-product is most likely to have an effect, but sometimes it is altogether unclear.

In order to generate the half-fraction design with r factors, we wrote down a full design in k – 1 factors, and added a column by aliasing the last factor with the (k – 1)th-order cross-product term. In the case of the quarter-fraction, we wrote down a full design in k – 2 factors, and then aliased each of the two remaining factors with (k – 3)th-order cross-products. We could have chosen to alias H5 with the product H1H2H3H4, and H6 with H2H3H4H5. If we had, then the identity relation would have been:

I = H1H2H3H4H5 = H2H3H4H5H6 = H1H6.

Thus, the single factor terms H1 and H6 would be aliased each other, and we would only be able to include one of them in the model.

Just as in the case of the half-fraction, there are multiple quarter-fraction designs associated with any pair of cross-product terms. In other words, we could have used any combination of ±H1H2H3H5 and ±H2H3H4H6 to generate the columns for H5 and H6. Regardless of which of the four combinations we chose, the resulting design would be ResIV, and the same models could be fit using any of the four resulting designs.

Smaller Fractions and Resolution III Designs

In a ResIII design, all single-factor terms are aliased with K – k = 3 – 1 = 2, or second-order cross-products. That means our model will only include single-factor terms (again invoking Mr. Ockham's rule). Such designs and models are mostly useful when there are so many factors (>5) and the EAS

Table 6.7 An eighth-fraction design in seven factors

Run	W1	W2	W3	W4	W5 = W1 W2	W6 = W2W3	W7 = W3W4
1	−1	−1	−1	−1	1	1	1
2	−1	−1	−1	1	1	1	−1
3	−1	−1	1	−1	1	−1	−1
4	−1	−1	1	1	1	−1	1
5	−1	1	−1	−1	−1	−1	1
6	−1	1	−1	1	−1	−1	−1
7	−1	1	1	−1	−1	1	−1
8	−1	1	1	1	−1	1	1
9	1	−1	−1	−1	−1	1	1
10	1	−1	−1	1	−1	1	−1
11	1	−1	1	−1	−1	−1	−1
12	1	−1	1	1	−1	−1	1
13	1	1	−1	−1	1	−1	1
14	1	1	−1	1	1	−1	−1
15	1	1	1	−1	1	1	−1
16	1	1	1	1	1	1	1

must choose a smaller subset (at least initially) with which to experiment. The ResIII designs are like heats in a race; they really meant as a means of eliminating factors. Consider a 2^{7-3} design, with seven factors in $2^{7-3} = 2^4 = 16$ runs. To generate the design, write down a full 2^4 design in H1, H2, H3, and H4, and then compute columns for H5, H6, and H7 by aliasing these single-factor terms with second-order cross-products. Table 6.7 shows the design.

The identity is:

$$I = H1H2H5 = H2H3H6 = H3H4H7$$

That is, there are $p = 3$ generators in this 2^{-3} fraction of a 2^7 deign. As in the case of the quarter fraction, the products of these identity elements are also identity elements. There are three pairs of products and 1 three-way product. So, the complete definition of the identity is:

$$I = H1H2H5 = H2H3H6 = H3H4H7 = H1H3H5H6 = H1H2H3H4H5H7 = H2H4H6H7 = H1H4H5H6H7$$

The aliases for the single-factor terms are:

$$H1*I = H2H5 = H1H2H3H6 = H1H3H4H7 = H3H5H6 = H2H3H4H5H7 = H1H2H4H6H7 = H4H5H6H7$$
$$H2*I = H1H5 = H3H6 = H2H3H4H7 = H1H2H3H5H6 = H1H3H4H5H7 = H4H6H7 = H1H2H4H5H6H7$$
$$H3*I = H1H2H3H5 = H2H6 = H2H4H7 = H1H5H6 = H1H2H4H5H7 = H2H3H4H6H7 = H1H3H4H5H6H7$$
$$H4*I = H1H2H4H5 = H2H4H6 = H3H7 = H1H3H4H5H6 = H1H2H3H5H7 = H2H6H7 = H1H5H6H7$$
$$H5*I = H1H2 = H2H3H5H6 = H3H4H5H7 = H1H3H6 = H1H2H3H4H7 = H2H4H5H6H7 = H1H4H6H7$$
$$H6*I = H1H2H5H6 = H2H3 = H3H4H6H7 = H1H3H5 = H1H2H3H4H5H6H7 = H2H4H7 = H1H4H5H7$$
$$H7*I = H1H2H5H7 = H2H3H6H7 = H3H4 = H1H3H5H6H7 = H1H2H3H4H5 = H2H4H6 = H1H4H5H6$$

All the single-factor terms are aliased with a second-order cross-product. The model should be simply:

$$y = \beta_0 + \sum_{i=1}^{r} \beta_i H_i + \varepsilon$$

Some Terms and Some Generalities

The collection of terms used as identities are referred to as the *defining relation* of the fractional design. Each particular term is called a *word* in the defining relation. The size of a word is the number of factors comprising the word. In the case of half fractions, there is always exactly one word in a defining relation. In order to obtain the highest resolution, the *word* for the half-fraction should be comprised of k factors. In the case of the quarter fraction, there are always three words in the defining relation, one for each of the $p = 2$ *generator words*, and the product of those two *words*.

It is generally desirable to maximize the resolution of a fractional design. The resolution is equal to the smallest *word* in the *defining relation*. For a half-fraction, there is always exactly one word in the defining relation. In a quarter-fraction, there are always three. In an eighth-fraction design with k factors, there are

$$\binom{3}{1}+\binom{3}{2}+\binom{3}{3}$$

words in the *defining relation*. In general, for a 2^{-p} fraction, the number of *words* in the *defining relation* is:

$$\binom{p}{1}+\binom{p}{2}+\binom{p}{3}+\cdots+\binom{p}{p}=2^{p}-1$$

We usually will choose to alias p of the single-factor terms such that the smallest word in the defining relation will be as large as possible. The resolution will depend on the number of runs (i.e., the size of the fraction) and factors, namely $N=2^{k-p}$ and k. Unfortunately, there is not in general a unique mapping from k and p to a *defining relation*. The smallest word in the defining relation will be the product of the two words with the greatest number of common factors.

Recall that there are $2^{p}\ 2^{-p}$ fractions of any design for a given defining relation. There is no particular statistical benefit or detriment in choosing any of these 2^{p} fractions. The fraction in which all the words have positive signs is called the *principle fraction*.

Several statistical software packages will generate the runs for fractional designs.

Figure 6.1 shows a screen shot of Minitab 16 menu selections for generating a fractional factorial design.

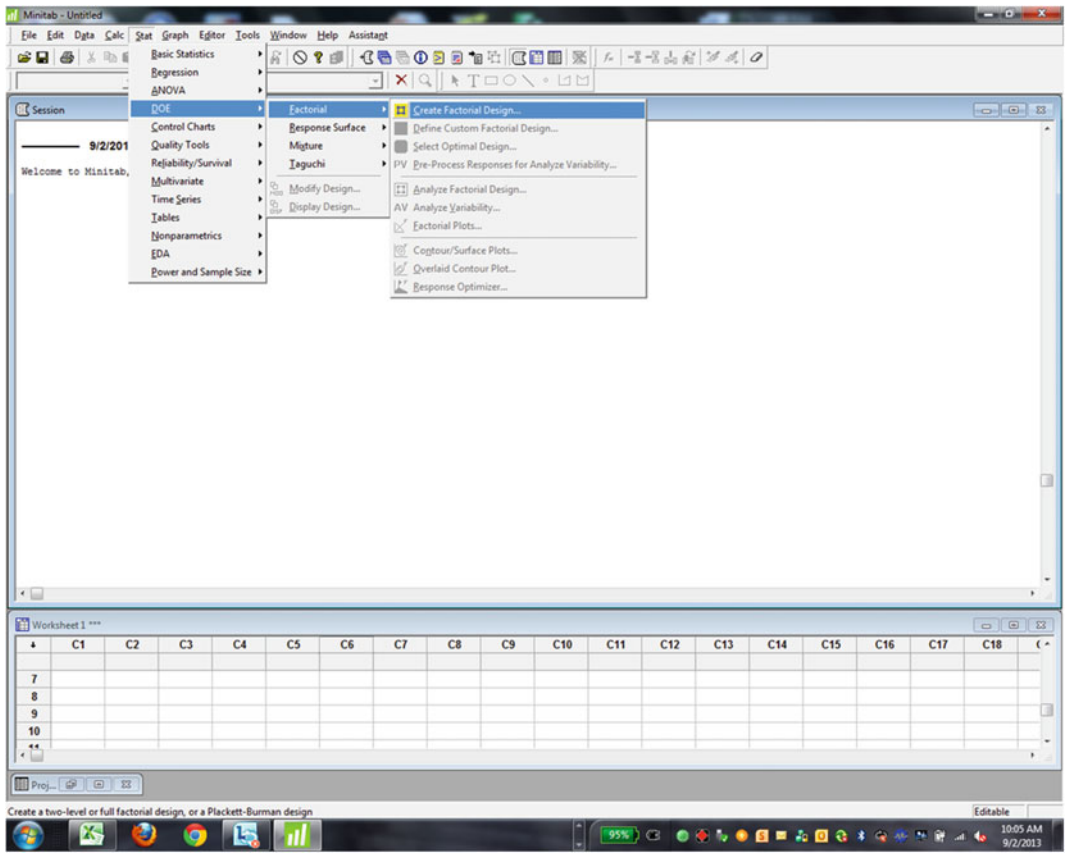


Fig. 6.1 Minitab 16 screen shot—creating a fractional factorial design—Part 1

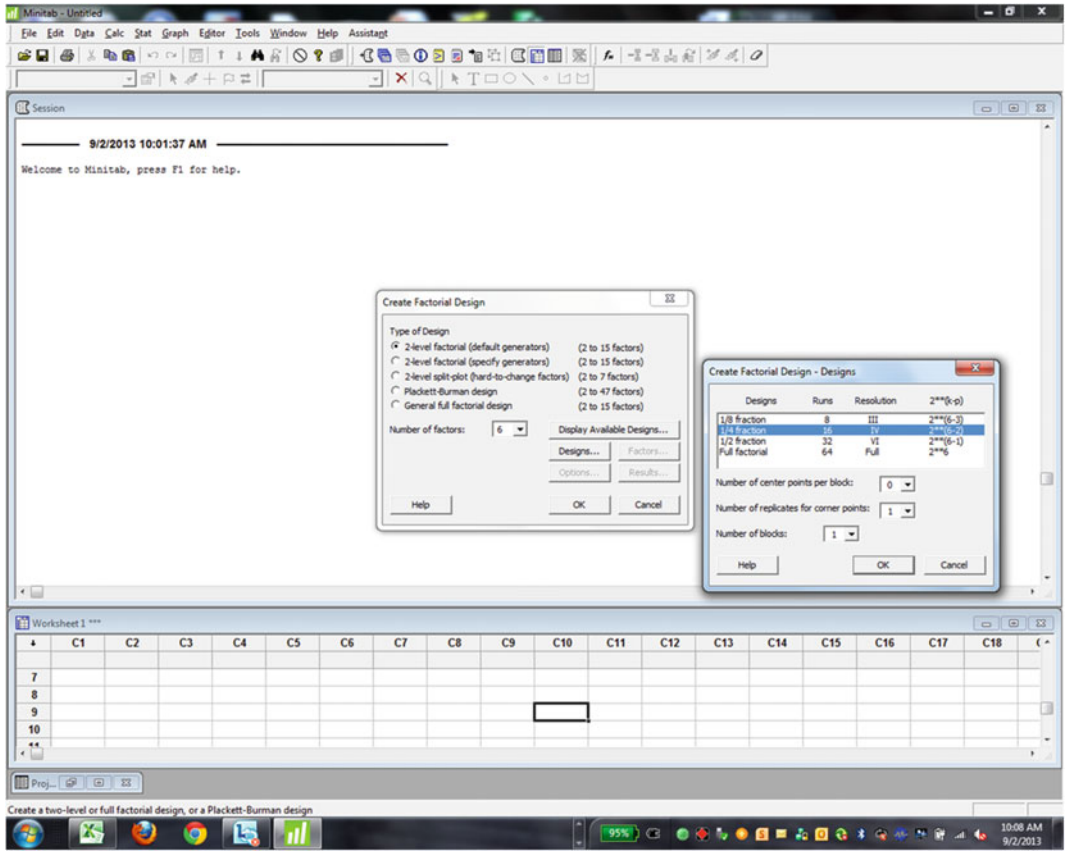


Fig. 6.2 Minitab 16 screen shot—creating a fractional factorial design—Part 2

Figure 6.2 shows the next Minitab screens in the design process.

We have asked Minitab to select the design with the default generator words, that we have six factors, and that we want a 16-run design. Note that in the main "Create Factorial Design" window, the user may select the "Factors" button to provide names for the factors, and even select codes for the low and high levels of each factor. The default coding is Helmert. The "Options" button allows the user to control how the design columns will be presented, and order in which the run rows will appear in the worksheet.

Table 6.8 shows the Session Window output, which describes the design and its alias structure.

The factors are named A-F, even though the user may have chosen more meaningful names. The "+" signs are used in the way we have previously used "=" in describing aliases.

Figure 6.3 is a screen shot of the Minitab worksheet. In the worksheet, the factor names given by the user appear as column headings. The columns labeled CenterPt and Blocks have no meaning in this context; their meaning will be apparent later in the text. The columns StdOrder and RunOrder are identical here. The user has the option of asking Minitab to provide an order in which the different runs should be executed. In many cases, the user may want to perform the runs in a random order, in order to mitigate the potential effects of order on the response.

A similar process for generating the design can be achieved with SAS 9.1 PROC FACTEX. Figure 6.4 shows the SAS code to generate the design.

The output from these statements is shown in Fig. 6.5.

This SAS procedure uses the symbol "0" to represent the complete alias structure for the identity, or defining relation.

Table 6.8 The 2^{6-2} fractional design—Minitab output

```
9/2/2013 10:01:37 AM

Welcome to Minitab, press F1 for help.

Fractional Factorial Design

Factors:   6   Base Design:         6, 16   Resolution:   IV
Runs:     16   Replicates:              1   Fraction:    1/4
Blocks:    1   Center pts (total):      0

Design Generators: E = ABC, F = BCD

Alias Structure

I + ABCE + ADEF + BCDF

A + BCE + DEF + ABCDF
B + ACE + CDF + ABDEF
C + ABE + BDF + ACDEF
D + AEF + BCF + ABCDE
E + ABC + ADF + BCDEF
F + ADE + BCD + ABCEF
AB + CE + ACDF + BDEF
AC + BE + ABDF + CDEF
AD + EF + ABCF + BCDE
AE + BC + DF + ABCDEF
AF + DE + ABCD + BCEF
BD + CF + ABEF + ACDE
BF + CD + ABDE + ACEF
ABD + ACF + BEF + CDE
ABF + ACD + BDE + CEF
```

The designs generated by Minitab and SAS are very similar, but not identical. However, both designs are ResIV. In the Minitab design, since H1H5 = H2H3 (AE = BC in Minitab's notation), only one of these two-way cross-products could be included in the model. In the SAS design, H1H5 is not aliased with H2H3, so both of those terms could be included in the model.

JMP, another software package, can also be used to generate design tables. Figure 6.6 shows the screens, together with the output table.

JMP refers to fractional factorials as "screening designs". It also adds a column in the output table that identifies the pattern of low and high levels for each run. For example, the pattern "-----" indicates that the run has all six factors set to their "low" (−1) level.

Fries and Hunter (1980) described another measure of assessing design quality in addition to resolution. They called it aberration. It is the number of words in the defining relation with the smallest size. It is possible for two designs to have the same resolution, yet one may be somewhat more desirable. That is, the fewer words in the defining relation with the smallest size, then lower the amount of aliasing. Hinkelmann and Kempthorne (2005) give an example of two 2^{8-3}, ResIV designs. The first, they call D1, has the defining relation:

I = H1H2H3H4H5 = H3H4H5H7 = H1H2H6H7 = H2H4H5H8 = H1H3H6H8 = H2H3H7H8 = H1H4H5H6H7H8

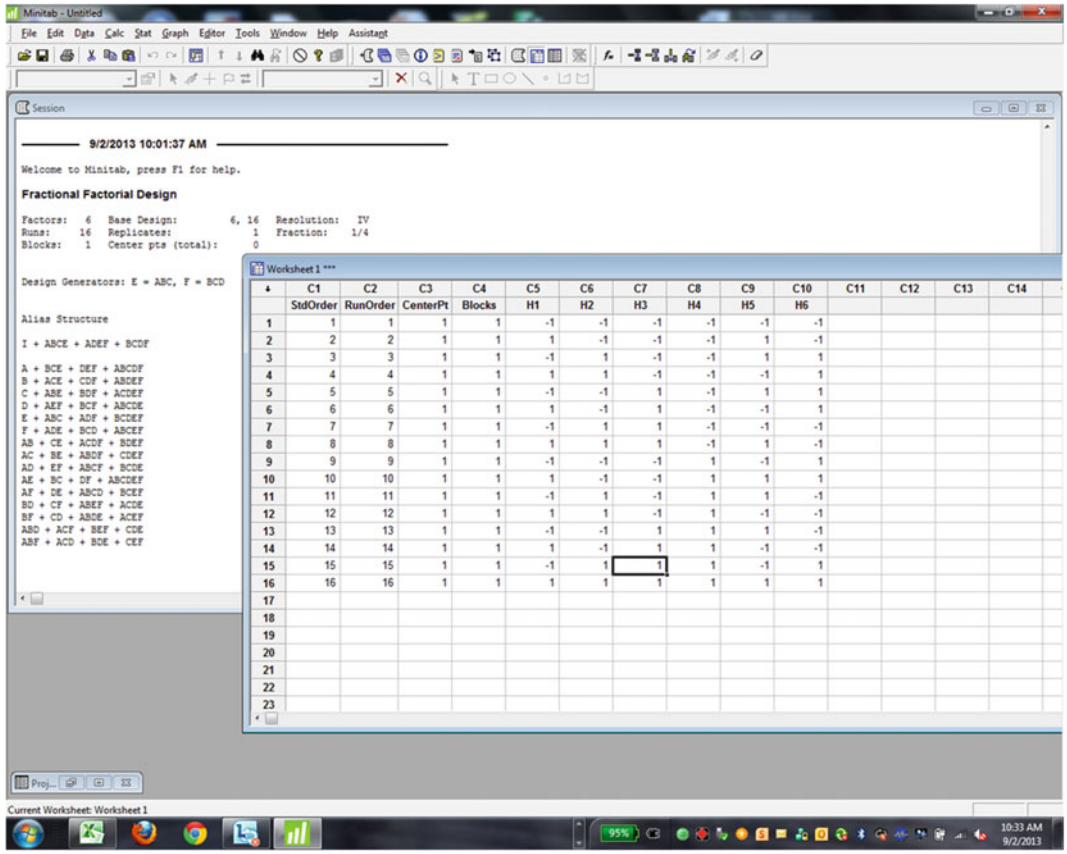


Fig. 6.3 The 2^{6-2} fractional design—Minitab worksheet

The other, D2, has the relation:

$$I = H3H4H5H6 = H1H2H4H5H7 = H1H2H3H6H7 = H1H2H3H5H8 = H3H4H7H8 = H5H6H7H8$$

Design D2 has only three words of size 4, but D1 has 5. Both designs are ResIV, but in some sense D2 may be more desirable. SAS PROC FACTEX can be told explicitly to provide the minimum aberration design for a given resolution. However, it is not necessarily true that the minimum aberration design is best. Usually, the default design for a given resolution and/or maximum number of runs will be perfectly sufficient. In the case of ResIII designs, usually the objective of "screening", or reducing the number of factors in which to experiment, the first-order model approximation is very coarse. Thus, concern over aberration may be of secondary or even lower priority relative to creating a design with a sufficiently low number of runs. In the case of ResV designs, minimizing aberration would result in reducing the numbers of four-way cross-products to which the two-way cross-products are aliased. Again, this is probably of considerably lesser importance than obtaining a ResV design. The biggest issue is for ResIV designs, since two-way cross-products are aliased with each other. However, minimizing aberration generally only affects the three-way and higher order cross-product terms, which are more likely to be excluded from the models fit to ResIV design data anyway. However, especially in ResIV designs, care must be taken to note which two-way cross-products are aliased, so that an appropriate model can be fit.

```
options nonumber nodate linesize=80 pagesize=54;
libname fact 'H:\Personal Data\Experimentation for Design &
Validation\Data & Analyses';
/* use lib fact to create an output dataset
   containing the design
*/

proc factex;

  factors H1 H2 H3 H4 H5 H6;
  size fraction=4;
  model res=4;
  examine design aliasing(6) confounding;
  output out=frac;
  title1 'Figure 6.4 - 2**(6-2) Res IV Design';
  run;

  data fact.f20130903_resIV_2_6_design;
    set frac;

      run;
```

Fig. 6.4 SAS Proc FACTEX code for generating the 2^{6-2} ResIV design

Blocking Effects

Blocking effects are factors that are not of particular interest to the designer, but are unavoidable constraints to making experimental runs. For example, raw material for making an injection-molded part may come in batches, or lots, and differences between lots may affect the value of the response variable. When all the runs cannot be completed using a single batch, or in other words, within a single block, then a decision must be made as to how to split the runs between two or more blocks. One method is to include the block variable as another factor, and in designing a fractional factorial, alias the block factor with some higher-order interaction (Montgomery 2001). For example, suppose the EAS wants to do a full four factor, two-level experiment. The coded factor names are H1, H2, H3, and H4. This is a 2^4 experiment, with 16 runs. The EAS cannot perform all 16 runs in a single day, and she is concerned that there may be some day-to-day differences. She lets a fifth factor, B1, equal to the product of the levels (rows) of the other four factors, B1 = H1H2H3H4. The factor B1 represents Day, and it is aliased with the four-way cross-product term. Table 6.9 shows the runs. The value of B1 determines on which day the run is to be conducted (e.g., B1 = −1 is for Day 1, and B1 = +1 is for Day 2). This method for assigning run to blocks works well for 2^{k-p} fractional designs if the number of blocks required to perform all the necessary runs is a power of 2. If not, then splitting the runs into multiple blocks is more complex. For more complicated blocking systems, the reader is directed to Cochran and Cox (1992).

Figure 6.5 - 2**(6-2) Res IV Design

The FACTEX Procedure

Design Points

Experiment Number	H1	H2	H3	H4	H5	H6
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
13	1	1	-1	-1	1	1
14	1	1	-1	1	-1	-1
15	1	1	1	-1	-1	-1
16	1	1	1	1	1	1

Table 6.8 - 2**(6-2) Res IV Design

The FACTEX Procedure

Factor Confounding Rules

H5 = H2*H3*H4
H6 = H1*H3*H4

Figure 6.5 - 2**(6-2) Res IV Design
The FACTEX Procedure

Aliasing Structure

0 = H1*H2*H5*H6 = H1*H3*H4*H6 = H2*H3*H4*H5
H1 = H2*H5*H6 = H3*H4*H6 = H1*H2*H3*H4*H5
H2 = H1*H5*H6 = H3*H4*H5 = H1*H2*H3*H4*H6
H3 = H1*H4*H6 = H2*H4*H5 = H1*H2*H3*H5*H6
H4 = H1*H3*H6 = H2*H3*H5 = H1*H2*H4*H5*H6
H5 = H1*H2*H6 = H2*H3*H4 = H1*H3*H4*H5*H6
H6 = H1*H2*H5 = H1*H3*H4 = H2*H3*H4*H5*H6
H1*H2 = H5*H6 = H1*H3*H4*H5 = H2*H3*H4*H6
H1*H3 = H4*H6 = H1*H2*H4*H5 = H2*H3*H5*H6
H1*H4 = H3*H6 = H1*H2*H3*H5 = H2*H4*H5*H6
H1*H5 = H2*H6 = H1*H2*H3*H4 = H3*H4*H5*H6
H1*H6 = H2*H5 = H3*H4 = H1*H2*H3*H4*H5*H6
H2*H3 = H4*H5 = H1*H2*H4*H6 = H1*H3*H5*H6
H2*H4 = H3*H5 = H1*H2*H3*H6 = H1*H4*H5*H6
H1*H2*H3 = H1*H4*H5 = H2*H4*H6 = H3*H5*H6
H1*H2*H4 = H1*H3*H5 = H2*H3*H6 = H4*H5*H6

Fig. 6.5 Proc FACTEX output

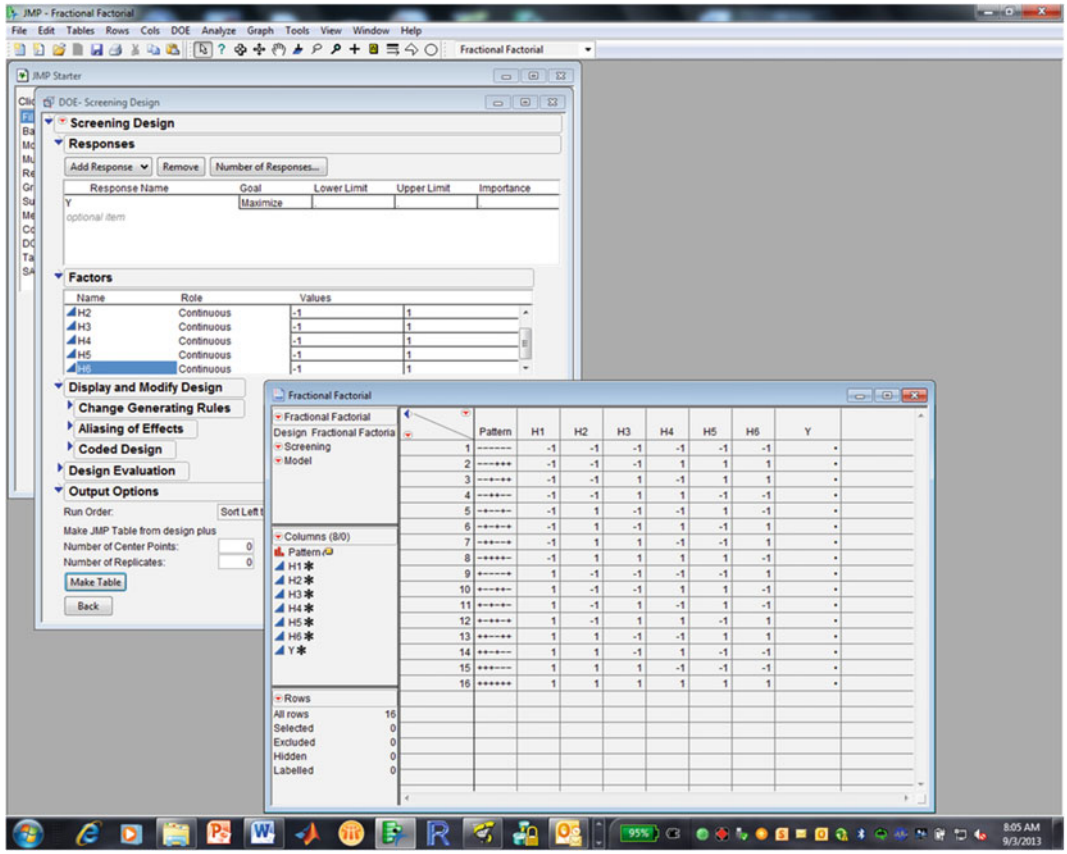


Fig. 6.6 JMP 2^{6-2} Res IV design

The Moral

The moral of the fractional factorial design story is:

1. Fractional designs can save money/time, but you will pay in terms of the information and level of approximating the response function you can achieve;
2. ResV designs provide a premium balance between level of approximation and economy; all the two-way cross-products are aliased with only the higher-order terms, which generally provide less impact on the approximation;
3. ResIII designs are really only useful to narrow down, in a rational fashion, the number of factors in which to experiment;
4. ResIV designs can be useful, especially if it is possible to use prior knowledge to determine which two-way cross-product terms are most likely to have an impact on the response variable(s);
5. The generation of these fractional designs is done in a manner that preserves balance and orthogonality, the two properties that allow for the optimal estimation of the model coefficients; the generation techniques also control the amount of information that is lost (resolution);
6. There are several software packages that will generate the runs for fractional designs.

Table 6.9 A 2^4 experiment in two blocks

Pattern	H1	H2	H3	H4	B1 = H1H2H3H4
−−−−	−1	−1	−1	−1	1
−−−+	−1	−1	−1	1	−1
−−+−	−1	−1	1	−1	−1
−−++	−1	−1	1	1	1
−+−−	−1	1	−1	−1	−1
−+−+	−1	1	−1	1	1
−++−	−1	1	1	−1	1
−+++	−1	1	1	1	−1
+−−−	1	−1	−1	−1	−1
+−−+	1	−1	−1	1	1
+−+−	1	−1	1	−1	1
+−++	1	−1	1	1	−1
++−−	1	1	−1	−1	1
++−+	1	1	−1	1	−1
+++−	1	1	1	−1	−1
++++	1	1	1	1	1

Table 6.10 Hypodermic needle variables—five factors

Variable	Factor	Units	Low (−1)	High (+1)
X1	Primary angle	degrees	14.5	15.5
X2	Secondary angle	degrees	9.5	10.5
X3	Side angle	degrees	4.5	5.5
X4	Primary length	mm	0.245	0.255
X5	Secondary length	mm	0.095	0.105

Examples

ResV

An EAS is trying to design an hypodermic needle. The needle point geometry has three angles, primary bevel angle (X1), secondary bevel angle (*X2*), and side bevel angle (X3). In addition to the angles, there are the primary bevel length (X4) and secondary bevel length (X5) factors. The EAS decides to perform a two-level experiment in the five factors. He can afford 16 runs, and he can perform $n = 3$ replicates for each run. Thus he decides to perform a 2^{5-1} half-fraction, which is ResV. The factors and their associated low and high levels, are given in Table 6.10. The runs, with Helmert-coded variables, are shown in Table 6.11.

The defining relation for this design has only one word:

I = H1H2JH3H4H5

Since it is a half-fraction, it has only one generator. We have chosen the principal fraction, having purposely aliased H5 with the four-way cross-product H5 = +H1H2H3H4.

Accordingly, each single factor term is aliased with no lower than a four-way cross-product, and each two-way cross-product is aliased with no lower than a three-way cross-product. Again, following Ockham, the model will be:

Table 6.11 2^{5-1} ResV design with coded levels

Run	Primary angle X1 H1	Secondary angle X2 H2	Side angle X3 H3	Primary length X4 H4	Secondary length X5 H5
1	−1	−1	−1	−1	1
2	−1	−1	−1	1	−1
3	−1	−1	1	−1	−1
4	−1	−1	1	1	1
5	−1	1	−1	−1	−1
6	−1	1	−1	1	1
7	−1	1	1	−1	1
8	−1	1	1	1	−1
9	1	−1	−1	−1	−1
10	1	−1	−1	1	1
11	1	−1	1	−1	1
12	1	−1	1	1	−1
13	1	1	−1	−1	1
14	1	1	−1	1	−1
15	1	1	1	−1	−1
16	1	1	1	1	1

$$y = \beta_0 + \sum_{i=1}^{r} \beta_i H_i + \sum_{i \neq j} \sum \gamma_{ij} H_i H_j + \varepsilon$$

The response variable for this needle design experiment is the penetration force generated using a fixture that penetrates a needle through simulated skin, driving the needle at a constant velocity. The force the needle imparts on the artificial skin membrane is measured continuously, from the time the needle contacts the membrane to the time after the needle tip has broken through and completely penetrated, and the peak force is recorded. Lower forces imply a better needle. Table 6.12 shows all the data for the 16 runs.

Figure 6.7 shows the output from JMP "Fit Model" function. Note that the p-values for testing the hypotheses that model coefficients are actually 0 are greater than 0.05 in most of the cases. In fact, the only terms with significant coefficients are H1 (or X1 = primary angle), H3 (or X3 = side angle), H5 (X5 = secondary length) and the interaction, or cross-product terms, H1*H4 and H2*H5. Although by themselves, H2 and H4 did not have significant coefficients, since they are factors in significant two-way cross-products, they are potentially important. In the spirit of parsimony, we will drop out those two-way cross-products from the model, and use predicted values from a reduced model:

$$y = \beta_0 + \beta_1 H_1 + \beta_2 H_2 + \beta_3 H_3 + \beta_4 H_4 + \beta_5 H_5 + \gamma_{14} H_1 H_4 + \gamma_{25} H_2 H_5 + \varepsilon$$

Figure 6.8 shows the JMP output for the reduced model. The adjusted R^2 was slightly increased in the reduced model

Figure 6.9 shows the R code and output for the same model.

Figure 6.10 shows the SAS code and output for the same reduced model.

When using Minitab to fit the same model, the user must create columns for the cross-products. Figure 6.11 illustrates via screen shot how such columns can be computed.

Figure 6.12 shows the Minitab output.

Table 6.12 2^{5-1} ResV needle design data

Run	H1	H2	H3	H4	H5	Peak force	Run	H1	H2	H3	H4	H5	Peak force
1	−1	−1	−1	−1	1	0.092	9	1	−1	−1	−1	−1	0.085
1	−1	−1	−1	−1	1	0.098	9	1	−1	−1	−1	−1	0.077
1	−1	−1	−1	−1	1	0.114	9	1	−1	−1	−1	−1	0.061
2	−1	−1	−1	1	−1	0.026	10	1	−1	−1	1	1	0.156
2	−1	−1	−1	1	−1	0.043	10	1	−1	−1	1	1	0.108
2	−1	−1	−1	1	−1	0.033	10	1	−1	−1	1	1	0.134
3	−1	−1	1	−1	−1	0.121	11	1	−1	1	−1	1	0.119
3	−1	−1	1	−1	−1	0.142	11	1	−1	1	−1	1	0.113
3	−1	−1	1	−1	−1	0.106	11	1	−1	1	−1	1	0.103
4	−1	−1	1	1	1	0.064	12	1	−1	1	1	−1	0.155
4	−1	−1	1	1	1	0.058	12	1	−1	1	1	−1	0.15
4	−1	−1	1	1	1	0.059	12	1	−1	1	1	−1	0.124
5	−1	1	−1	−1	−1	0.063	13	1	1	−1	−1	1	0.107
5	−1	1	−1	−1	−1	0.058	13	1	1	−1	−1	1	0.097
5	−1	1	−1	−1	−1	0.046	13	1	1	−1	−1	1	0.118
6	−1	1	−1	1	1	0.077	14	1	1	−1	1	−1	0.106
6	−1	1	−1	1	1	0.059	14	1	1	−1	1	−1	0.084
6	−1	1	−1	1	1	0.073	14	1	1	−1	1	−1	0.104
7	−1	1	1	−1	1	0.182	15	1	1	1	−1	−1	0.04
7	−1	1	1	−1	1	0.171	15	1	1	1	−1	−1	0.047
7	−1	1	1	−1	1	0.146	15	1	1	1	−1	−1	0.101
8	−1	1	1	1	−1	0.025	16	1	1	1	1	1	0.205
8	−1	1	1	1	−1	−0.003	16	1	1	1	1	1	0.202
8	−1	1	1	1	−1	0.03	16	1	1	1	1	1	0.203

ResIV

Suppose that instead of only having five input factors for needle design, the EAS also has an additional variable, namely lubrication quantity (X6). With six factors, even if they are limited to two levels, the total number of possible runs is $2^6 = 64$. The EAS can only afford to perform 16 of these 64 possible combinations. Thus, he requires a 2^{6-2}, or one-quarter fractional design. SAS Proc Factex can be used to generate such a design, as illustrated in Fig. 6.13.

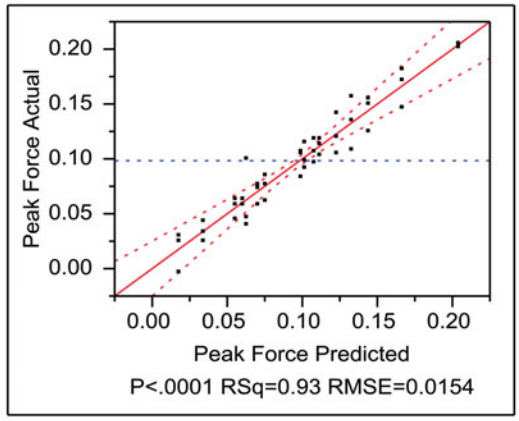
What's wrong with this design? Earlier we found that the only two-way cross-products that seemed to be significant were H1H4 and H2H5. In the above design, the cross-product H1H6 (primary angle * lubrication) is aliased with H2H5 (secondary angle * secondary bevel length). As we have no information on H6 or H1H6, we would rather not have it aliased with something we already know is significant. JMP allows us to alter the generators for the design and examine the alias structure. Figure 6.14 is a screen shot of the design window.

H1*H2 = H3*H6
H1*H3 = H2*H6 = H4*H5
H1*H4 = H3*H5
H1*H5 = H3*H4
H1*H6 = H2*H3
H2*H4 = H5*H6
H2*H5 = H4*H6

The generators for this design are:

H5 = H1H3H4
H6 = H1H2H3

Response Peak Force
Actual by Predicted Plot



Summary of Fit

RSquare	0.934704
RSquare Adj	**0.904097**
Root Mean Square Error	0.015431
Mean of Response	0.097524
Observations (or Sum Wgts)	48

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	15	0.10908229	0.007272	30.5385
Error	32	0.00762019	0.000238	**Prob > F**
C. Total	47	0.11670248		<.0001*

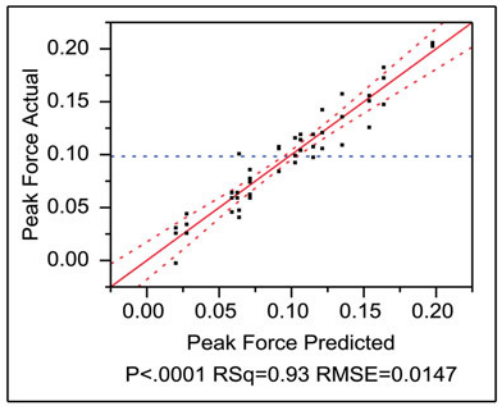
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	**0.0975242**	**0.002227**	**43.78**	**<.0001***
H1	**0.0191509**	**0.002227**	**8.60**	**<.0001***
H2	0.0000191	0.002227	0.01	0.9932
H3	**0.0133806**	**0.002227**	**6.01**	**<.0001***
H4	-0.002715	0.002227	-1.22	0.2318
H5	**0.0215434**	**0.002227**	**9.67**	**<.0001***
H1*H2	0.0011526	0.002227	0.52	0.6084
H1*H3	8.2933e-5	0.002227	0.04	0.9705
H1*H4	**0.0303754**	**0.002227**	**13.64**	**<.0001***
H1*H5	0.0005212	0.002227	0.23	0.8165
H2*H3	0.001496	0.002227	0.67	0.5066
H2*H4	0.0023208	0.002227	1.04	0.3052
H2*H5	**0.0176203**	**0.002227**	**7.91**	**<.0001***
H3*H4	-0.00218	0.002227	-0.98	0.3350
H3*H5	0.0028511	0.002227	1.28	0.2097
H4*H5	7.6315e-5	0.002227	0.03	0.9729

Fig. 6.7 JMP Fit Model output—ResV needle design experiment

Model

Response Peak Force
Actual by Predicted Plot



Summary of Fit

RSquare	0.925606
RSquare Adj	**0.912587**
Root Mean Square Error	0.014733
Mean of Response	0.097524
Observations (or Sum Wgts)	48

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	7	0.10802054	0.015432	71.0971
Error	40	0.00868193	0.000217	**Prob > F**
C. Total	47	0.11670248		<.0001*

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	8	0.00106174	0.000133	0.5573
Pure Error	32	0.00762019	0.000238	**Prob > F**
Total Error	40	0.00868193		0.8042
				Max RSq
				0.9347

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.0975242	0.002126	45.86	<.0001*
H1	0.0191509	0.002126	9.01	<.0001*
H2	0.0000191	0.002126	0.01	0.9929
H3	0.0133806	0.002126	6.29	<.0001*
H4	-0.002715	0.002126	-1.28	0.2090
H5	0.0215434	0.002126	10.13	<.0001*
H1*H4	0.0303754	0.002126	14.28	<.0001*
H2*H5	0.0176203	0.002126	8.29	<.0001*

Fig. 6.8 JMP "Fit Model" output—ResV needle design experiment—reduced model

```
df1 <- read.csv("20130909_Example_5_2_Needle.csv")

attach(df1)

first_order <- lm(Peak ~ H1 + H2 + H3 + H4 + H5 +
H1*H4 + H2*H5)

summary(first_order)

detach(df1)

Call:
lm(formula = Peak ~ H1 + H2 + H3 + H4 + H5 + H1 * H4 + H2 * H5)

Residuals:
     Min       1Q   Median       3Q      Max
-0.029708 -0.008313  0.001292  0.007208  0.037708

Coefficients:
             Estimate     Std. Error   t value   Pr(>|t|)
(Intercept)  9.754e-02 2.128e-03  45.830   < 2e-16 ***
H1           1.908e-02  2.128e-03   8.966    4.07e-11 ***
H2           3.093e-18  2.128e-03   0.000    1.000
H3           1.342e-02  2.128e-03   6.304    1.77e-07 ***
H4          -2.750e-03  2.128e-03  -1.292    0.204
H5           2.154e-02  2.128e-03  10.121    1.36e-12 ***
H1:H4        3.037e-02  2.128e-03  14.272   < 2e-16 ***
H2:H5        1.758e-02  2.128e-03   8.262    3.51e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01475 on 40 degrees of freedom
Multiple R-squared:  0.9254,    Adjusted R-squared:  0.9123
F-statistic: 70.88 on 7 and 40 DF,  p-value: < 2.2e-16
```

Fig. 6.9 R "lm" code and output—ResV needle design experiment—reduced model

In this design, H1H6 is not aliased with either H1H4 or H2H5. Furthermore, H2H6 is not aliased with either H1H4 or H2H5. Of course, H1H4 and H2H5 are not aliased with each other. There is, of course, no guarantee that the H4H6 cross-product is not significant, and it is aliased with H2H5, which we already believe is significant. Inasmuch as we already are operating under the constraints of

```
libname stuff 'H:\Personal Data\Experimentation for Design &
Validation\Data & Analyses';

data calc;

  set stuff.d20130909_example_5_2_needle;
/*************************************************************
*                                                            *
* variables: X1 = primary angle                              *
*            X2 = secondary angle                            *
*            X3 = side angle                                 *
*            X4 = primary length                             *
*            X5 = secondary length                           *
*                                                            *
*            Hi = Xi coded to (-1,+1)                        *
*                                                            *
*************************************************************/

  run;

proc means data=calc;
  var Peak;
  by Run;
  run;

proc glm data=calc;
  model Peak = H1 H2 H3 H4 H5
               H1*H4 H2*H5/est;
  run;
```

```
The SAS System      08:59 Thursday, September 12, 2013   6

                                The GLM Procedure

                        General Form of Estimable Functions

                Effect          Coefficients

                Intercept       L1

                H1              L2

                H2              L3

                H3              L4

                H4              L5

                H5              L6

                H1*H4           L7

                H2*H5           L8
The SAS System      08:59 Thursday, September 12, 2013   7
```

Fig. 6.10 SAS "Proc GLM" code and output—ResV needle design experiment—reduced model

The GLM Procedure

Dependent Variable: Peak Peak

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	0.10802054	0.01543151	71.10	<.0001
Error	40	0.00868193	0.00021705		
Corrected Total	47	0.11670248			

R-Square	Coeff Var	Root MSE	Peak Mean
0.925606	15.10657	0.014733	0.097524

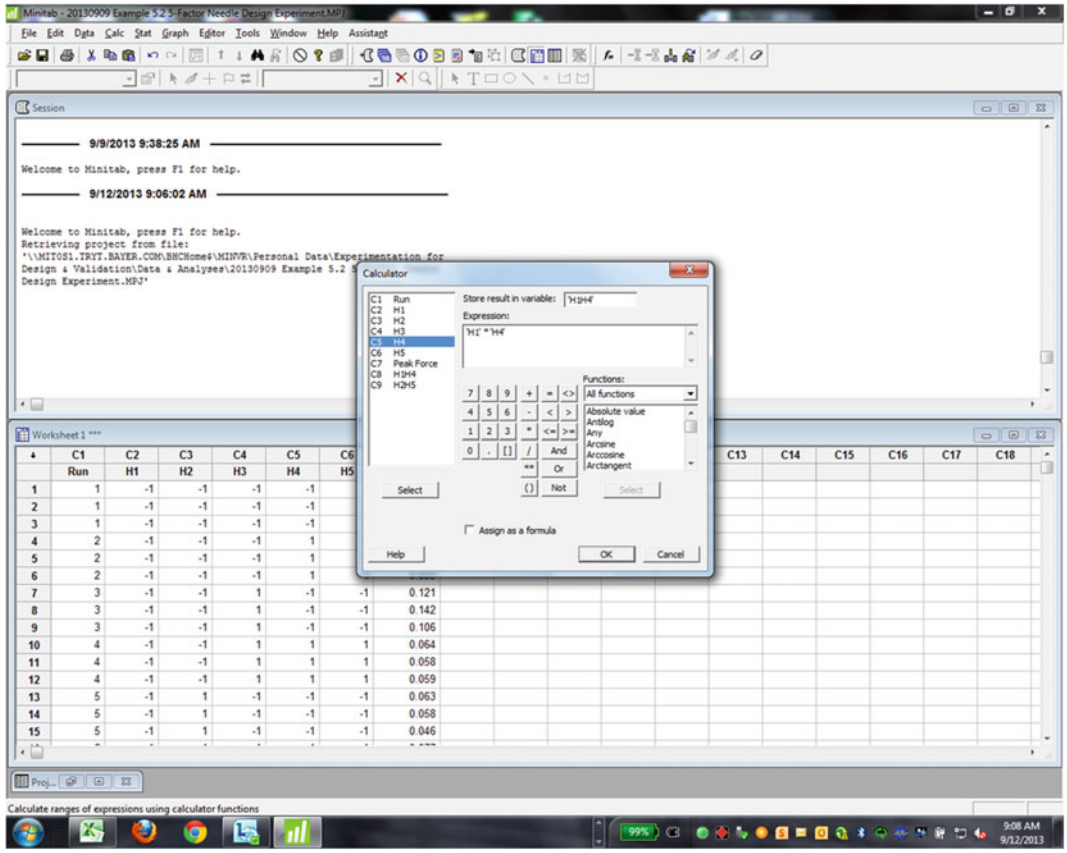
Source	DF	Type I SS	Mean Square	F Value	Pr > F
H1	1	0.01760429	0.01760429	81.11	<.0001
H2	1	0.00000002	0.00000002	0.00	0.9929
H3	1	0.00859391	0.00859391	39.59	<.0001
H4	1	0.00035383	0.00035383	1.63	0.2090
H5	1	0.02227776	0.02227776	102.64	<.0001
H1*H4	1	0.04428787	0.04428787	204.05	<.0001
H2*H5	1	0.01490287	0.01490287	68.66	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
H1	1	0.01760429	0.01760429	81.11	<.0001
H2	1	0.00000002	0.00000002	0.00	0.9929
H3	1	0.00859391	0.00859391	39.59	<.0001
H4	1	0.00035383	0.00035383	1.63	0.2090
H5	1	0.02227776	0.02227776	102.64	<.0001
H1*H4	1	0.04428787	0.04428787	204.05	<.0001
H2*H5	1	0.01490287	0.01490287	68.66	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.0975241683	0.00212646	45.86	<.0001
H1	0.0191508759	0.00212646	9.01	<.0001
H2	0.0000190995	0.00212646	0.01	0.9929
H3	0.0133805777	0.00212646	6.29	<.0001
H4	-.0027150337	0.00212646	-1.28	0.2090
H5	0.0215434438	0.00212646	10.13	<.0001
H1*H4	0.0303753825	0.00212646	14.28	<.0001
H2*H5	0.0176203436	0.00212646	8.29	<.0001

Fig. 6.10 (continued)

only allowing 16 runs, and that we also realize we are at best approximating the relationship between peak force and the six input variables, we must recognize that such an experiment and modeling exercise is flawed. However, the flaws are outweighed by obtaining at least an ability to approximate. Furthermore, once a model is constructed, it is not chiseled in stone; a new model may be constructed, and potentially additional experimentation may be used to adjudicate which cross-products are truly active.



	C1	C2	C3	C4	C5	C6	C7
	Run	H1	H2	H3	H4	H5	
1	1	-1	-1	-1	-1		
2	1	-1	-1	-1	-1		
3	1	-1	-1	-1	-1		
4	2	-1	-1	-1	1		
5	2	-1	-1	-1	1		
6	2	-1	-1	-1	1		
7	3	-1	-1	1	-1	-1	0.121
8	3	-1	-1	1	-1	-1	0.142
9	3	-1	-1	1	-1	-1	0.106
10	4	-1	-1	1	1	1	0.064
11	4	-1	-1	1	1	1	0.058
12	4	-1	-1	1	1	1	0.059
13	5	-1	1	-1	-1	-1	0.063
14	5	-1	1	-1	-1	-1	0.058
15	5	-1	1	-1	-1	-1	0.046

Fig. 6.11 Minitab screen shot—computing a cross-product column

Based on the above alias structure, and the prior experience, we might consider fitting the model:

$$y = \beta_0 + \beta_1 H_1 + \beta_2 H_2 + \beta_3 H_3 + \beta_4 H_4 + \beta_5 H_5 + \beta_6 H_6 + \gamma_{14} H_1 H_4 + \gamma_{25} H_2 H_5 + \gamma_{16} H_1 H_6 + \gamma_{26} H_2 H_6 + \varepsilon$$

Figure 6.15 shows the JMP output from fitting the model. As in the five-factor experiment, H1 (coded version of X1), H3, and H5 are all significant single-factor terms, and H2 and H4 are not significant. In addition, H6 (coded for X6 = lubrication quantity) is a significant single-factor term. The two-way cross-products, H1H4 and H2H5 are still significant, and so is H1H6. H2H6, however is not significant. As in the five factor model, we will keep H2 and H4 in the model, since they are part of significant two-way cross-products, but it appears that we could reduce the model by excluding the H2H6 cross-product. Figure 6.16 shows the JMP output for the reduced model. Although the estimates of the coefficients for terms included in the reduced model did not change, the adjusted R^2 actually increased, and the p-values for the coefficients changed slightly.

By way of illustration, Fig. 6.17 shows the SAS program and output for fitting the same reduced model.

```
Welcome to Minitab, press F1 for help.
Retrieving project from file:
'H:\\Personal Data\Experimentation for
Design & Validation\Data & Analyses\20130909 Example 5.2 5-Factor Needle
Design Experiment.MPJ'
```

Regression Analysis: Peak Force versus H1, H2, H3, H4, H5, H1H4, H2H5

```
The regression equation is
Peak Force = 0.0975 + 0.0191 H1 - 0.00000 H2 + 0.0134 H3 - 0.00275 H4
             + 0.0215 H5 + 0.0304 H1H4 + 0.0176 H2H5


Predictor        Coef   SE Coef      T      P
Constant     0.097542  0.002128  45.83  0.000
H1           0.019083  0.002128   8.97  0.000
H2          -0.000000  0.002128  -0.00  1.000
H3           0.013417  0.002128   6.30  0.000
H4          -0.002750  0.002128  -1.29  0.204
H5           0.021542  0.002128  10.12  0.000
H1H4         0.030375  0.002128  14.27  0.000
H2H5         0.017583  0.002128   8.26  0.000


S = 0.0147454   R-Sq = 92.5%   R-Sq(adj) = 91.2%


Analysis of Variance

Source           DF        SS        MS      F      P
Regression        7  0.107885  0.015412  70.88  0.000
Residual Error   40  0.008697  0.000217
Total            47  0.116582
```

Fig. 6.12 Minitab "Regression" output—ResV needle design experiment—reduced model

Res III

An EAS is designing a simple RLC circuit with impressed voltage of the form:

$$V(t) = V_0 \cos(\omega t)$$

She implements the circuit on an electronics "breadboard", in order to decide on parameter settings by experimentation. Table 6.13 shows the input parameters and the ranges of values over which she intends to experiment.

The variables Ih(0) and Ihdot(0) represent the initial conditions for the associated homogeneous equation. The output response variable of interest is the current over a 100 ms period. The question is which parameters are most important in terms of affecting this current. Even though the equation for an RLC circuit is well known: $L\ddot{Q}(t) + R\dot{Q}(t) + \frac{1}{C}Q(t) = V_0 \cos(\omega t)$ and $I(t) = \frac{dQ(t)}{dt}$, it is not clear which inputs have the largest effect on current. Furthermore, due to thermal and other sources of noise, the current is not deterministically known. The EAS realizes that a full, $2^7 = 128$ run factorial experiment is not realistically possible. In fact, even though she will use a breadboard "prototype" to

Figure 6.13 - 2**(6-2) Res IV Design with minabs

The FACTEX Procedure

Design Points

Experiment Number	H1	H2	H3	H4	H5	H6
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
13	1	1	-1	-1	1	1
14	1	1	-1	1	-1	-1
15	1	1	1	-1	-1	-1
16	1	1	1	1	1	1

Figure 6.13 - 2**(6-2) Res IV Design with minabs

The FACTEX Procedure

Factor Confounding Rules

H5 = H2*H3*H4
H6 = H1*H3*H4

Figure 6.13 - 2**(6-2) Res IV Design with minabs

The FACTEX Procedure

Aliasing Structure

0 = H1*H2*H5*H6 = H1*H3*H4*H6 = H2*H3*H4*H5
H1 = H2*H5*H6 = H3*H4*H6 = H1*H2*H3*H4*H5
H2 = H1*H5*H6 = H3*H4*H5 = H1*H2*H3*H4*H6
H3 = H1*H4*H6 = H2*H4*H5 = H1*H2*H3*H5*H6
H4 = H1*H3*H6 = H2*H3*H5 = H1*H2*H4*H5*H6
H5 = H1*H2*H6 = H2*H3*H4 = H1*H3*H4*H5*H6
H6 = H1*H2*H5 = H1*H3*H4 = H2*H3*H4*H5*H6
H1*H2 = H5*H6 = H1*H3*H4*H5 = H2*H3*H4*H6
H1*H3 = H4*H6 = H1*H2*H4*H5 = H2*H3*H5*H6
H1*H4 = H3*H6 = H1*H2*H3*H5 = H2*H4*H5*H6
H1*H5 = H2*H6 = H1*H2*H3*H4 = H3*H4*H5*H6
H1*H6 = H2*H5 = H3*H4 = H1*H2*H3*H4*H5*H6
H2*H3 = H4*H5 = H1*H2*H4*H6 = H1*H3*H5*H6
H2*H4 = H3*H5 = H1*H2*H3*H6 = H1*H4*H5*H6
H1*H2*H3 = H1*H4*H5 = H2*H4*H6 = H3*H5*H6
H1*H2*H4 = H1*H3*H5 = H2*H3*H6 = H4*H5*H6

Fig. 6.13 SAS Factex output—2^{6-2} fractional factorial

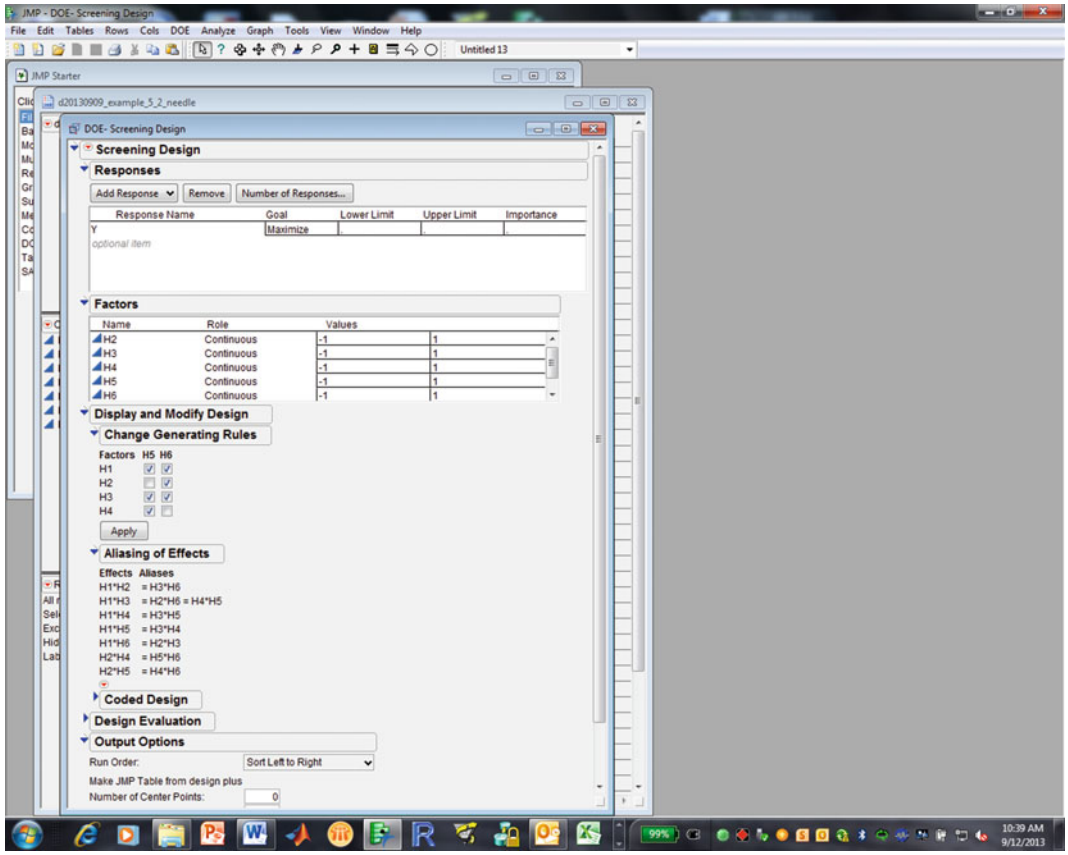


Fig. 6.14 JMP screening design window—altering the design generators

make measurements, she can only afford the time to make eight runs, with two replicate current measurements for each run. The data are given in Table 6.14.

The column labeled "Pattern" indicates the levels of each of the seven factors for the run ("−" means low level, "+" means high level).

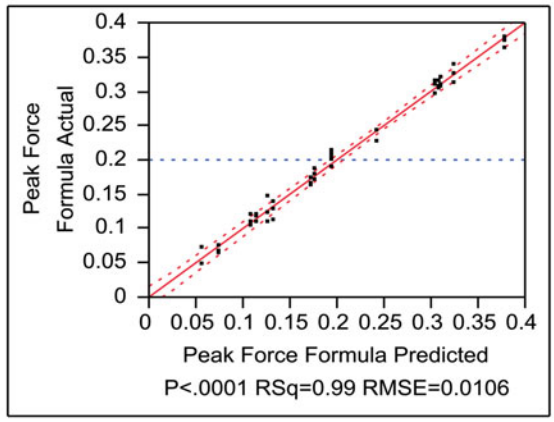
Since this was a Res III experiment, no cross-product terms can be included in the approximating formula. Figure 6.18 shows the JMP output with coefficient estimates for each of the single-factor terms, and the intercept.

All the factors are significant ($p < 0.05$). The factors L, C, omega, and V_0 have coefficients that are an order of magnitude or more greater than those of the other factors. Thus, it appears that these factors have the most influence, so that further experimentation might involve these four factors only, with the others set to a constant value.

In this case, the natural values for the levels of the factors were used to fit the model (i.e., estimate the coefficients). One of the disadvantages of using natural units is scale. That is, each factor is expressed in different units, so interpreting relative differences in the magnitude of the model coefficients might be difficult. Figure 6.19 shows the output of fitting the same model, but using Helmert-coded factors.

By using coded units, the coefficients are now on a common scale. The EAS can now see that inductance, L, is not as influential as it had appeared when the factors were expressed in natural units. Capacitance was the most influential, followed by the phase angle. The initial charge, Ihdot(0), appears to be more important than initial current, Ih(0), which was not the conclusion the EAS might draw using the natural units.

Actual by Predicted Plot



Summary of Fit

RSquare	0.990599
RSquare Adj	**0.988058**
Root Mean Square Error	0.010647
Mean of Response	0.200598
Observations (or Sum Wgts)	48

Analysis of Variance

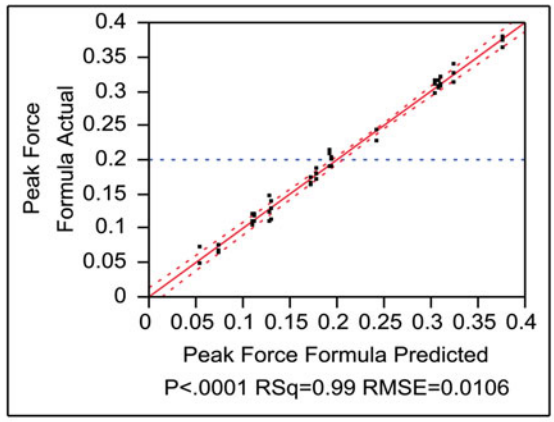
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	10	0.44194408	0.044194	389.8746
Error	37	0.00419415	0.000113	**Prob > F**
C. Total	47	0.44613823		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.2005975	0.001537	130.53	<.0001*
H1	0.0198903	0.001537	12.94	<.0001*
H2	0.0002717	0.001537	0.18	0.8606
H3	0.011411	0.001537	7.43	<.0001*
H4	0.001042	0.001537	0.68	0.5019
H5	0.0216945	0.001537	14.12	<.0001*
H6	-0.078184	0.001537	-50.88	<.0001*
H1*H4	0.0303036	0.001537	19.72	<.0001*
H2*H5	0.0167656	0.001537	10.91	<.0001*
H1*H6	-0.029938	0.001537	-19.48	<.0001*
H2*H6	0.0010424	0.001537	0.68	0.5018

Fig. 6.15 JMP output from fitting model to 2^{6-2} ResIV needle design experiment

Response Peak Force Formula
Actual by Predicted Plot



Summary of Fit

RSquare	0.990482
RSquare Adj	**0.988228**
Root Mean Square Error	0.010571
Mean of Response	0.200598
Observations (or Sum Wgts)	48

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	9	0.44189192	0.049099	439.3855
Error	38	0.00424631	0.000112	**Prob > F**
C. Total	47	0.44613823		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.2005975	0.001526	131.47	<.0001*
H1	0.0198903	0.001526	13.04	<.0001*
H2	0.0002717	0.001526	0.18	0.8596
H3	0.011411	0.001526	7.48	<.0001*
H4	0.001042	0.001526	0.68	0.4988
H5	0.0216945	0.001526	14.22	<.0001*
H6	-0.078184	0.001526	-51.24	<.0001*
H1*H4	0.0303036	0.001526	19.86	<.0001*
H2*H5	0.0167656	0.001526	10.99	<.0001*
H1*H6	-0.029938	0.001526	-19.62	<.0001*

Fig. 6.16 JMP output from fitting reduced model to 2^{6-2} ResIV needle design experiment

```
libname stuff 'H:\Personal Data\Experimentation for Design &
Validation\Data & Analyses';

data calc;

  set stuff.d20130917_example_5_3_needle6;
/*************************************************************
*                                                            *
* variables: X1 = primary angle                              *
*            X2 = secondary angle                            *
*            X3 = side angle                                 *
*            X4 = primary length                             *
*            X5 = secondary length                           *
*            X6 = lubrication quantity                       *
*                                                            *
*            Hi = Xi coded to (-1,+1)                        *
*                                                            *
*************************************************************/

  run;


proc glm data=calc;
  model Peak = H1 H2 H3 H4 H5 H6
               H1*H4 H2*H5 H1*H6/est;
  run;
```

```
The SAS System       07:36 Tuesday, September 17, 2013   6

                                The GLM Procedure

                        General Form of Estimable Functions

                            Effect        Coefficients

                            Intercept     L1

                            H1            L2

                            H2            L3

                            H3            L4

                            H4            L5

                            H5            L6

                            H6            L7

                            H1*H4         L8

                            H2*H5         L9

                            H1*H6         L10
Dependent Variable: Peak   Peak

                                    Sum of
        Source                DF    Squares     Mean Square    F Value    Pr > F
```

Fig. 6.17 SAS code and output from fitting reduced model to 2^{6-2} ResIV needle design experiment

Model	9	0.44189192	0.04909910	439.39	<.0001
Error	38	0.00424631	0.00011174		
Corrected Total	47	0.44613823			

R-Square	Coeff Var	Root MSE	Peak Mean
0.990482	5.269730	0.010571	0.200598

Source	DF	Type I SS	Mean Square	F Value	Pr > F
H1	1	0.01898999	0.01898999	169.94	<.0001
H2	1	0.00000354	0.00000354	0.03	0.8596
H3	1	0.00625016	0.00625016	55.93	<.0001
H4	1	0.00005212	0.00005212	0.47	0.4988
H5	1	0.02259123	0.02259123	202.17	<.0001
H6	1	0.29341381	0.29341381	2625.75	<.0001
H1*H4	1	0.04407866	0.04407866	394.46	<.0001
H2*H5	1	0.01349209	0.01349209	120.74	<.0001
H1*H6	1	0.04302032	0.04302032	384.99	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
H1	1	0.01898999	0.01898999	169.94	<.0001
H2	1	0.00000354	0.00000354	0.03	0.8596
H3	1	0.00625016	0.00625016	55.93	<.0001
H4	1	0.00005212	0.00005212	0.47	0.4988
H5	1	0.02259123	0.02259123	202.17	<.0001
H6	1	0.29341381	0.29341381	2625.75	<.0001
H1*H4	1	0.04407866	0.04407866	394.46	<.0001
H2*H5	1	0.01349209	0.01349209	120.74	<.0001
H1*H6	1	0.04302032	0.04302032	384.99	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.2005975432	0.00152578	131.47	<.0001
H1	0.0198903195	0.00152578	13.04	<.0001
H2	0.0002717242	0.00152578	0.18	0.8596
H3	0.0114110357	0.00152578	7.48	<.0001
H4	0.0010420384	0.00152578	0.68	0.4988
H5	0.0216944836	0.00152578	14.22	<.0001
H6	-.0781843184	0.00152578	-51.24	<.0001
H1*H4	0.0303035538	0.00152578	19.86	<.0001
H2*H5	0.0167655957	0.00152578	10.99	<.0001
H1*H6	-.0299375457	0.00152578	-19.62	<.0001

Fig. 6.17 (continued)

Table 6.13 Circuit parameters

Factor	Symbol	Low (−1)	High (+1)	Units
Inductance	L	0.2	0.5	henry
Resistance	R	300	500	ohm
Capacitance	C	0.0001	0.0005	farad
Frequency	ω	0.5	1	radian
Initial voltage	V_0	0.5	1	volt
Initial current	Ih(0)	1	2	amp
Initial charge	Ihdot(0)	1	2	coulomb

Table 6.14 Data for 2^{7-4} Res III RLC experiment

Pattern	L	R	C	omega	V_0	Ih(0)	Ihdot(0)	I meas.
−−−−+++	0.2	300	0.0001	0.5	1	2	2	1.40E−05
−−−−+++	0.2	300	0.0001	0.5	1	2	2	1.39E−05
−−++−−+	0.2	300	0.0005	1	0.5	1	2	1.55E−04
−−++−−+	0.2	300	0.0005	1	0.5	1	2	1.56E−04
−+−+−+−	0.2	500	0.0001	1	0.5	2	1	2.66E−05
−+−+−+−	0.2	500	0.0001	1	0.5	2	1	2.80E−05
−++−+−−	0.2	500	0.0005	0.5	1	1	1	9.33E−05
−++−+−−	0.2	500	0.0005	0.5	1	1	1	9.68E−05
+−−++−−	0.5	300	0.0001	1	1	1	1	5.30E−05
+−−++−−	0.5	300	0.0001	1	1	1	1	5.47E−05
+−+−−+−	0.5	300	0.0005	0.5	0.5	2	1	3.98E−05
+−+−−+−	0.5	300	0.0005	0.5	0.5	2	1	4.19E−05
++−−−−+	0.5	500	0.0001	0.5	0.5	1	2	6.70E−06
++−−−−+	0.5	500	0.0001	0.5	0.5	1	2	8.08E−06
+++++++	0.5	500	0.0005	1	1	2	2	3.40E−04
+++++++	0.5	500	0.0005	1	1	2	2	3.41E−04

A Special ResIII Design: Plackett-Burman

If there are many factors that potentially have importance, ResIII designs are useful to narrowing down the experimental space to the "vital few". In particular, if number of factors, k, is such that $k + 1$ is a multiple of 4, and not a power of 2 (e.g., $k = 11$, 19, or 23), then a special class of ResIII designs, called Plackett-Burman Designs (PBD, Montgomery 2001), with number of runs equal to $N = k + 1$, can be useful. To generate the design matrix for a PBD, the EAS must begin with a sort of "seed" column, which depends on the number of factors. The column will have k rows. The seed column is assigned to the first factor, H1. The column for the second factor, H2, is created by assigning the value in the kth row of the previous column to the first row of the new column, and then the value of the first row of the previous column becomes the value of the second row, the second row of the previous column becomes the value in the third row, etc., until k rows are completed. Once k columns are generated, add a single row to the bottom of the matrix, where this row in every column is set to the low level (−1) of the corresponding factor. Table 6.15 shows a list of "seed" columns for $k = 11$, 19, 23, 27. Note that a "−" implies the Helmert-coded level of −1, and "+" the level +1.

As an example, for $k = 11$ factors, the PBD would look like Table 6.16.

The Plackett-Burman design is balanced, with equal numbers of runs with low (−1) and high (+1) levels of each factor. Furthermore, the columns are all orthogonal to each other. One criticism of the PBD is that the alias structure is very difficult to determine (Hinkelmann and Kempthorne 2005). However, if the EAS has decided to use a ResIII design, simply knowing that all interactions are unresolvable, or completely aliased with each other and the main effects, may be sufficient. For a more complete set of seed columns, see Cochran and Cox (1992).

Response I meas.
Summary of Fit

RSquare	0.999929
RSquare Adj	0.999868
Root Mean Square Error	1.24e-6
Mean of Response	9.178e-5
Observations (or Sum Wgts)	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	7	1.74407e-7	2.4915e-8	16206.52
Error	8	1.2299e-11	1.537e-12	Prob > F
C. Total	15	1.74419e-7		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.000568	2.415e-6	-235.0	<.0001*
L	0.0001253	2.067e-6	60.65	<.0001*
R	2.5719e-7	3.1e-9	82.97	<.0001*
C	0.3307891	0.00155	213.43	<.0001*
omega	0.0002099	1.24e-6	169.29	<.0001*
V0	0.000136	1.24e-6	109.69	<.0001*
Ih(0)	0.0000276	6.2e-7	44.53	<.0001*
Ihdot(0)	7.506e-5	6.2e-7	121.07	<.0001*

Fig. 6.18 Analysis output for 2^{7-4} Res III RLC experiment

Response I meas.
Summary of Fit

RSquare	0.999929
RSquare Adj	0.999868
Root Mean Square Error	1.24e-6
Mean of Response	9.178e-5
Observations (or Sum Wgts)	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	7	1.74407e-7	2.4915e-8	16206.52
Error	8	1.2299e-11	1.537e-12	Prob > F
C. Total	15	1.74419e-7		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.0000918	3.1e-7	296.09	<.0001*
L(0.2,0.5)	0.0000188	3.1e-7	60.65	<.0001*
R(300,500)	0.0000257	3.1e-7	82.97	<.0001*
C(0.0001,0.0005)	0.0000662	3.1e-7	213.43	<.0001*
omega(0.5,1)	0.0000525	3.1e-7	169.29	<.0001*
V0(0.5,1)	0.0000340	3.1e-7	109.69	<.0001*
Ih(0)(1,2)	0.0000138	3.1e-7	44.53	<.0001*
Ihdot(0)(1,2)	0.0000375	3.1e-7	121.07	<.0001*

Fig. 6.19 Analysis output for 2^{7-4} Res III RLC experiment—Helmert-coded factors

Table 6.15 Some "Seed" columns for selected Plackett-Burman designs

Row	k = 11	k = 19	k = 23	k = 27
1	+	+	+	+
2	+	+	+	−
3	−	−	+	+
4	+	−	+	+
5	+	+	+	+
6	+	+	−	+
7	−	−	+	−
8	−	+	−	−
9	−	−	+	−
10	+	+	+	−
11	−	−	−	+
12		+	−	−
13		−	+	−
14		−	+	−
15		−	−	+
16		−	−	−
17		+	+	−
18		+	−	+
19		−	+	+
20			−	+
21			−	−
22			−	+
23			−	−
24				+
25				+
26				−
27				+

Table 6.16 A k = 11 Plackett-Burman design

Run	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11
1	1	−1	1	−1	−1	−1	1	1	1	−1	1
2	1	1	−1	1	−1	−1	−1	1	1	1	−1
3	−1	1	1	−1	1	−1	−1	−1	1	1	1
4	1	−1	1	1	−1	1	−1	−1	−1	1	1
5	1	1	−1	1	1	−1	1	−1	−1	−1	1
6	1	1	1	−1	1	1	−1	1	−1	−1	−1
7	−1	1	1	1	−1	1	1	−1	1	−1	−1
8	−1	−1	1	1	1	−1	1	1	−1	1	−1
9	−1	−1	−1	1	1	1	−1	1	1	−1	1
10	1	−1	−1	−1	1	1	1	−1	1	1	−1
11	−1	1	−1	−1	−1	1	1	1	−1	1	1
12	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1

Key Points

- Fractional Factorial experiments allow the experimenter to run a subset (fraction) of the total number of combinations of levels of all factors, and still be able to fit a first-order model.
- The drawback of fractional factorials is called aliasing; aliasing means that some effects are not resolvable, because they are computed identically to some other effects. The level of fractional replication dictates the degree of aliasing.

- There are classes of the degree of aliasing, called Resolution; Res V allows the experimenter to include all two-way cross-product terms in the model; Res IV allows for including some two-way cross-products; Res III only allows for first-order terms (no cross-products).
- Helmert coding transforms the range of the input factors to the interval [−1,+1]. This transformation is unitless and scale-free.
- A center point is a set of factor values where each factor is set at the midpoint of its experimental range. Results from center points can be useful in detecting the presence of second-order effects.

Exercises and Questions

1. Generate a ResIII experiment with 11 factors. What model could be fit to the data from this design if all 11 factors are included? What is the defining relation for the design?
2. Describe a situation where a ResIV design would be appropriate.
3. How, and under what circumstances, would you convince stakeholders that a ResV design is worth its cost?

Chapter 7
Higher Order Approximations

Two-level factorial or fractional factorial experimental designs are used for obtaining a first-order approximation to the response function. They are particularly useful for selecting a smaller subset of potential input factors with which to formulate a better approximation equation. In this chapter, we will discuss some classes of experimental designs useful for fitting second-order (Quadratic) approximating equations.

A two-level experiment cannot support the inclusion of second-order terms in a model. If a center point is included in a two-level design, a plot of model residuals against model predicted values may indicate the need for second-order terms. If the residual plot shows that at the predicted response for the center point, the residuals are appreciably higher or lower than they are at the other predicted values, then it is likely that a second-order effect exists, and a second-order model is required for good predictability. Consider a full two-level, three-factor experiment with replication at the center point, together with the observed response, the predicted values, and the residuals, shown in Table 7.1. A first-order model was fit to these data:

$$Y = \beta_0 + \beta_1 H1 + \beta_2 H2 + \beta_3 H3 + \beta_{12} H1{*}H2 + \beta_{13} H1{*}H3 + \beta_{23} H2{*}H3 + \varepsilon$$

The residual plot is shown in Fig. 7.1. Note that the residuals at the three replicates of the center point run are appreciably greater than the residuals at the other runs. This indicates that at least one factor has an associated higher-order effect on the response.

A Brief Digression: Residuals, Heteroscedasticity, and Normality

A plot of residuals against predicted values can help identify "outlier" points, which may have unusual response values due to unintended or uncontrolled effects. When examining a residual plot, the EAS should ask whether there were any extenuating conditions that might be the cause of an unusually high or low residual value. If, in the case of center points, all the residuals at the center seem to be higher or lower than all the other residuals, it is reasonable to guess that higher-order effects exist. In fact, if the residuals at any particular point seem to "group" together and depart from the values of residuals at other points, higher-order effects that are not accounted for in the model may exist. A simple method for deciding whether a residual is "unusually" large in magnitude, is to normalize the residuals, by dividing them by the standard deviation of the residuals. The normalized residuals are said to be "Studentized" (Draper and Smith 1998). The idea is that if residuals are

S.A. Pardo, *Empirical Modeling and Data Analysis for Engineers and Applied Scientists*,
DOI 10.1007/978-3-319-32768-6_7

Table 7.1 A two-level, three-factor experiment with a center point

Run	H1	H2	H3	Y	Predicted Y	Residual Y
1	−1	−1	−1	−10.15	−9.25	−0.90
2	−1	−1	1	−8.51	−6.94	−1.57
3	−1	1	−1	−5.26	−3.70	−1.57
4	−1	1	1	−1.61	−0.71	−0.90
5	1	−1	−1	−10.26	−8.70	−1.57
6	1	−1	1	0.73	1.62	−0.90
7	1	1	−1	−4.96	−4.07	−0.90
8	1	1	1	5.37	6.94	−1.57
9	0	0	0	1.32	−3.10	4.42
9	0	0	0	−0.68	−3.10	2.42
9	0	0	0	−0.10	−3.10	3.00

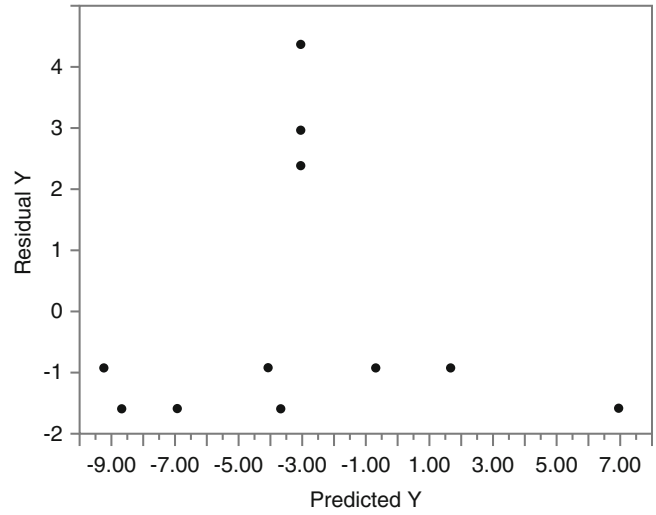


Fig. 7.1 Residual plot for the two-level, three-factor experiment with a center point

distributed as normal with mean 0 and some unknown standard deviation, then the Studentized residuals will be approximately t-distributed, with $n-1$ degrees of freedom. Thus, any Studentized residual that fell outside, say, a 99 % range for a t-distributed random variable with $n-1$ degrees of freedom ($t_{0.005}$, $t_{0.995}$), would be somewhat suspect. Furthermore, if more than about 1 % of residuals fell outside this range, the model might be suspect. In general, whenever there are residuals that seem appreciably higher or lower than the majority, the EAS should investigate possible "assignable causes".

Residuals plotted against predicted values can also reveal a lack of homogeneity of noise variance. If the plot looks like a funnel, with the small end of the funnel at the lower predicted values, and the larger end near the larger predicted values, then this is an indication that noise variance increases proportionally to the value of the response, a common form of heteroscedasticity. Figure 7.2 illustrates this phenomenon.

If this is the case, then it may be helpful to make a transformation of the response. The model can be fit to the transformed response, and the tests of significance for the coefficients may be more correct. A widely-used family of transformations is the Box-Cox transformation (Box et al. 1978):

$$Y_{trans} = \frac{{Y_{orig}}^{\lambda} - 1}{\lambda {Y_{orig}}^{\lambda-1}}$$

The value of λ is chosen to minimize the variance of residuals. A special limiting case when $\lambda \to 0$ is $Y_{trans} = \ln(Y_{orig})$. The Box-Cox transformation can only be applied when the response, Y_{orig}, is

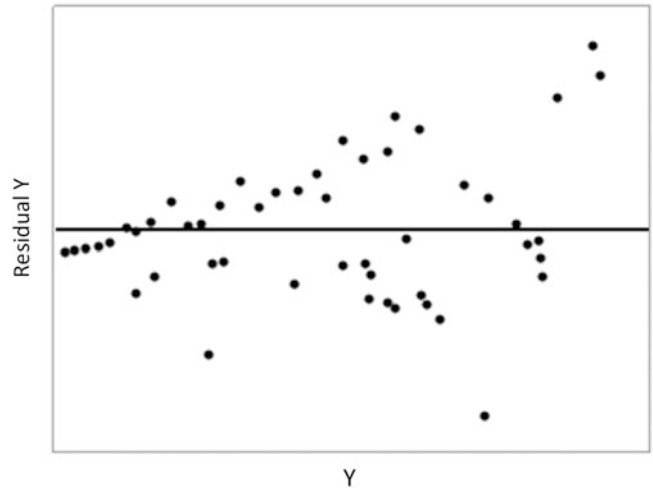


Fig. 7.2 Residual plot illustrating proportional noise variance

positive. In general, any obvious pattern in a residual vs. predicted plot usually indicates either higher-order effects not included in the model, or non-homogeneous noise variance.

Another assumption required for p-values to be valid is the normality of errors, or noise. One way to assess the validity of the normality assumption is to perform one of a number of tests using the residuals. The Shapiro-Wilk test and the Kolmogorov-Smirnov test (Conover 1999) are two such tests. Figure 7.3 shows the output of the distribution-fitting function in JMP. The **Prob** < **W** is the p-value for the Shapiro-Wilk test. If the p-value is below the pre-specified threshold (usually 0.05), then reject the null hypothesis that the data come from a normal population. If the residuals from a model fail a test for normality, there are a number of possible remedies. The Box-Cox transformations can sometimes induce normality. Replacing the observations with their relative ranks (i.e., the rank of the smallest observation is 1, and the rank of the largest is n) can be used to induce symmetry in the distribution. The Johnson family of transformations (Johnson 1949) can also be useful for inducing normality. Perhaps the most common is the S_U transformation (U stands for unbounded):

$$z = \gamma + \delta \sin h^{-1}((y - \xi)/\lambda)$$

The parameters, γ, δ, ξ, and λ can be estimated from data.

Another Johnson transformation is called S_B (B for bounded), and is given by:

$$z = \gamma + \delta \ln((y - \xi)/(\xi + \lambda - y)) \text{ for } \xi < y < \xi + \lambda$$

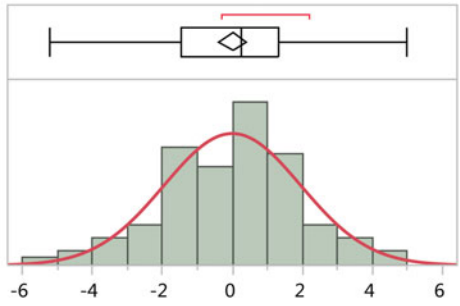
The new variable will have approximately a normal distribution with mean 0 and standard deviation 1 (called a "standard normal"), under the null hypotheses that none of the factors have any relationship to the response. Minitab will perform the estimation and provide the transformed data. Figure 7.4 shows the Minitab output transforming data column Y.

Note that Minitab uses the Anderson-Darling goodness-of-fit test, a test similar to the Shapiro-Wilk test. Also, Minitab decided to use the S_B transformation.

The histograms of the original data and the transformed data are given in Fig. 7.5.

Distributions

Residual Data



Normal(-7e-17,1.95786)

Quantiles

100.0%	maximum	4.99249
99.5%		4.99249
97.5%		3.87533
90.0%		2.18291
75.0%	quartile	1.34319
50.0%	median	0.24814
25.0%	quartile	-1.4752
10.0%		-2.7011
2.5%		-4.1095
0.5%		-5.23
0.0%	minimum	-5.23

Fitted Normal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	-7.11e-17	-0.388481	0.388481
Dispersion	σ	1.9578554	1.7190104	2.2743921

-2log(Likelihood) = 417.15764118077

Goodness-of-Fit Test

Shapiro-Wilk W Test

W	Prob<W
0.990752	0.7253

Note: Ho = The data is from the Normal distribution. Small p-values reject Ho.

Fig. 7.3 Histogram with normal density fit and Shapiro-Wilk test

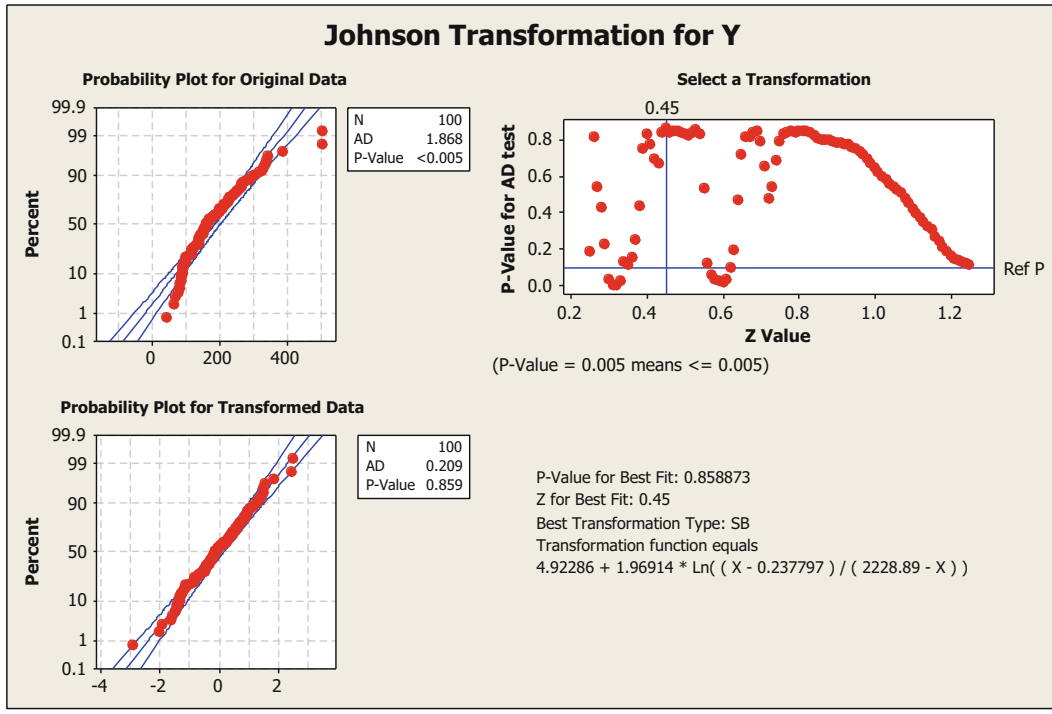


Fig. 7.4 Johnson SU transformation

Back to Second-Order Designs

By restricting the factor levels to only two possible values, the only approximation equation we can fit is first-order. In order to fit a second-order equation, we will need three values. If we had k potential input factors, and we generated an experiment in which each factor was varied with three levels, the number of runs would be 3^k. Even with a relatively small number of factors, the number or experimental runs would soon become large and probably too expensive. While fractional designs for three-level factors are quite possible, the aliasing structure of such designs is quite complex, and can be difficult at best to interpret. A center point (all factors set to their middle value, which corresponds to coded variables set to 0) would provide a third value. However, since it is identically the same point in the design for all the factors, the quadratic terms for each factor would be aliased with each other. We require a means for including a squared term for each factor uniquely in the approximating model. We will therefore only talk about two classes of experimental designs that will allow fitting a second-order approximating equation. Those two are called Central Composite Designs (CCDs) and Box-Behnken Designs (BBDs). Both of these classes of designs can be fairly economical, and both will allow the EAS to fit a second-order approximating equation.

Rotatability

In the two-level experiments, we were concerned with orthogonality and balance. For designs used to fit second-order approximations, another characteristic is important, namely rotatability. Rotatability is the characteristic that the standard error of predicted values would be the same for all points

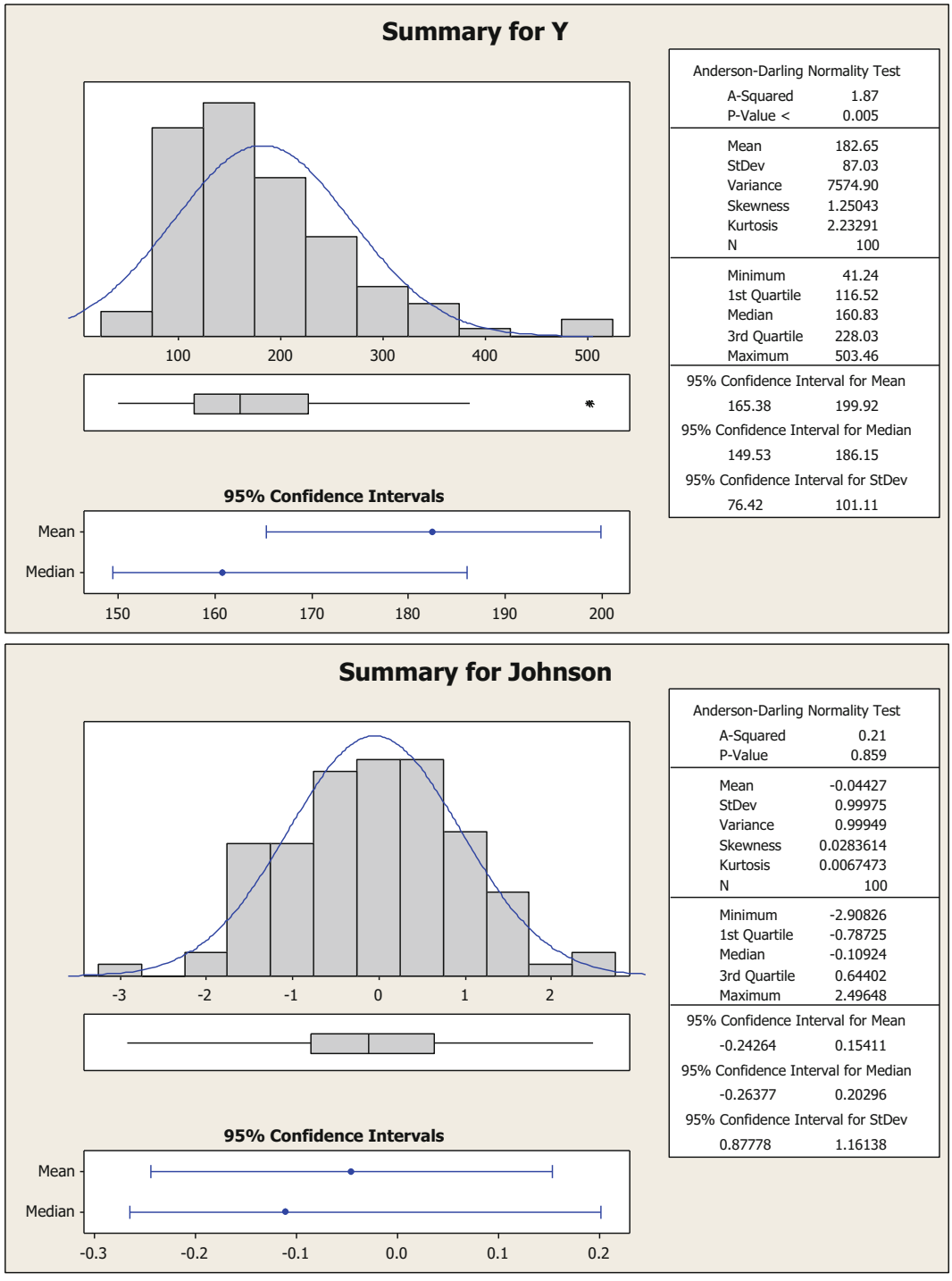


Fig. 7.5 Histograms of raw and Johnson-Transformed data

(experimental conditions) equidistant from the center point. The standard error for the predicted values are given by:

$$SE(\widehat{y}|\boldsymbol{h}_0) = s\sqrt{\boldsymbol{h}_0'\left[\boldsymbol{H}'\boldsymbol{H}\right]^{-1}\boldsymbol{h}_0}$$

Where $\boldsymbol{H}$ is a matrix of experimental runs, where entries are in Helmert-coded units. The lower case $\boldsymbol{h_0}$ represents a particular combination of values for the terms, which can be thought of as a point in the "term" space.

The CCD and BBD families are all rotatable.

CCD

A central composite design is a 2^{k-p} fractional factorial with two other conditions added: center points and axial points. Center points were discussed earlier. They are conditions where all the input factors are set to their "middle" value, i.e., the center of the range in which the experiment is being conducted. In some cases, where replicates are expensive or in some way difficult to obtain, center points may be the only condition in which replication is performed. Axial points are points that go beyond the "low" and "high" values set for experimentation, and allow for unique estimation of coefficients for quadratic terms in each factor. There are two axial conditions associated with each factor. If H_i represents the coded values for factor X_i, then the coded axial points would be where all the other H_j ($j \neq i$) variables are set equal to 0, and H_i is set to $\pm\sqrt{r}$, $|r| > 1$. Usually, r is set equal to k, the number of factors.

As an example, consider a two-factor experiment. The two coded factors are H1 and H2. The value -1 represents a "low" value for the corresponding factor in natural units, and +1 represents a "high" value. The runs in the CCD that correspond to those of the two-level design are referred to as "corner" points. The set of corner points for the two factors are given in Table 7.2.

The center point run is where H1 = H2 = 0. The axial points are where

H1 $= \pm\sqrt{r} = \pm\sqrt{2} \approx \pm 1.414$, H2 $= \pm\sqrt{r} = \pm\sqrt{2} \approx \pm 1.414$. Table 7.3 shows all the runs for this CCD.

Table 7.2 Corner point runs for a two-factor experiment

Run	H1	H2
1	−1	−1
2	−1	+1
3	+1	−1
4	+1	+1

Table 7.3 All runs for the two-factor CCD

Run	H1	H2
1	−1	−1
2	−1	+1
3	+1	−1
4	+1	+1
5	0	0
6	−1.414	0
7	+1.414	0
8	0	−1.414
9	0	+1.414

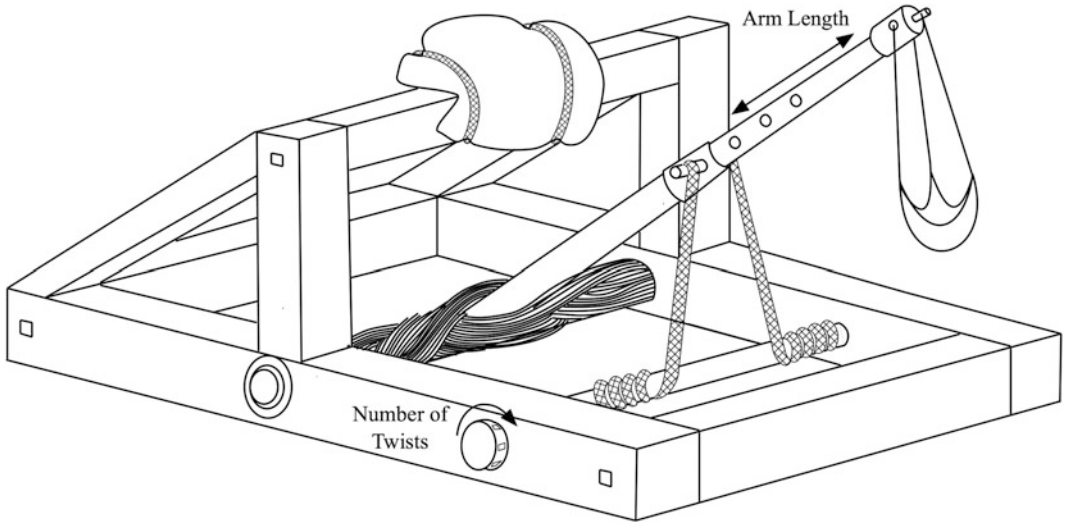


Fig. 7.6 The mangonal

Clearly, if for example, Run 6 requires that the coded variable H1 be set equal to -1.414, then H1 $= -1$ cannot be the code for the lowest POSSIBLE value of X1. Rather, H1 $= -1$ represents a low value for X1, but not the lowest possible value. The Helmert coding formula:

$$H = \frac{X - m}{\frac{1}{2}(x^+ - x^-)}$$

Is still valid, but x^- and x^+ no longer represent the most extreme values for the uncoded (natural units) factor. In fact, x^- and x^+ must be set such that

$$\max X \leq m + \frac{1}{2}R\sqrt{k} \text{ and } \min X \geq m - \frac{1}{2}R\sqrt{k}$$

where $R = x^+ - x^-$ and $m = \frac{x^+ + x^-}{2}$.

A Medieval EAS, Geoffrey Llewellyn, the Yeoman, is trying to decide how to set up a new kind of catapult, called a mangonal. The mangonal works by torsion. The throwing arm is attached to a twisted rope that provides tension. The arm is pulled down by another line, tightening the torsion rope. When the pulling line is released, the torsion rope unwinds, propelling the arm forward. A crossbar stops the arm, and the projectile (usually a rock) is released into ballistic flight. The crossbar angle governs the release angle for the projectile. Figure 7.6 shows a picture of a mangonal. Unbelievably, Geoffrey actually discovered all of the laws of motion attributed to Isaac Newton, who was born about 250 years after Geoffrey had faced his catapult problem. Some say that Sir Isaac actually discovered an old palimpsest, which had all of Geoffrey's notes covered by a portrait of the Earl of Phlegmingham, Geoffrey's Liege Lord. It turned out that the Earl was having a little war with his next-door neighbor, who had an imposing castle with lots of archers. The Earl asked Geoffrey to figure out how to deploy this new mangonal in an optimal fashion. Geoffrey figures that he and his crew must inflict some serious damage to the castle without getting shishkababed by enemy arrows.

Using the equations of ballistic flight, Geoffrey computed the altitude of the projectile after t units of time in flight:

$$y(t) = -\frac{1}{2}gt^2 + (v_0 \sin(\theta))t$$

g = gravitational constant (~9.8 m/s^2)
θ = launch angle
t = time (s)
v_0 = initial velocity (m/s)

The distance the projectile would travel after t units of time would be

$$x(t) = (v_0 \cos(\theta))t$$

Geoffrey wants the projectile to hit the window of the castle's "keep" (the most protected part of the castle) while he and his crew are out of range from the archers on the castle walls. He soon discovers that the arrows cannot fly more than100 m from the castle wall. The keep is approximately 25 m away from the wall, and spies have told Geoffrey that the window to the keep is approximately 9 m high with its center 40 m from the ground. The wall of the castle is only 15 m high. So, Geoff decides that his catapult must be placed 125 m away from the target, he wants the projectile to achieve an height of 40 m after it has traveled the 125 m. In other words, he wants to have t_f, the flight time to be such that:

$$y(t_f) = -\frac{1}{2}gt_f^2 + (v_0 \sin(\theta))t_f = 40$$

and

$$x(t_f) = (v_0 \cos(\theta))t_f = 125$$

These are of course two equations in three unknowns, t_f, θ, and v_0. This is further complicated by the fact that while the launch angle, θ, can be set by altering the position of the crossbar on the mangonal, the initial velocity, v_0, cannot easily be set. Geoffrey knows that v_0 is mostly a function of the number of twists in the launching rope, but this function is at best elusive, and pretty much unknown.

Geoffrey, being fairly adept at algebra, realizes that he can solve the $x(t)$ equation for t_f, and then substitute the solution in the $y(t)$ equation. This solving and substitution process gives him:

$$t_f = \frac{125}{v_0 \cos(\theta)}$$

$$y(v_0, \theta) = -\frac{1}{2}g\left(\frac{125}{v_0}\right)^2 sec^2(\theta) + 125\tan(\theta) - 40 = 0$$

Undaunted, Geoffrey decides to make one change of variables:

$$\eta_0 = \frac{1}{v_0}$$

So that the equation is now:

$$y(\eta_0, \theta) = -\frac{125^2}{2}g\eta_0^2 sec^2(\theta) + 125\tan(\theta) - 40 = 0$$

Geoffrey thinks about setting the launch angle, θ, equal to 45°. If he did, then his equation would be simply:

Table 7.4 Experimental runs for catapult experiment—natural units

Run	Rotations	Arm
1	10	2
2	10	3
3	15	2
4	15	3
5	9.0	2.5
6	16.0	2.5
7	12.5	1.8
8	12.5	3.2
9	12.5	2.5
10	12.5	2.5

Table 7.5 Experimental runs for catapult experiment—coded units

Run	H1	H2
1	−1	−1
2	−1	1
3	1	−1
4	1	1
5	−1.414	0.0
6	1.414	0.0
7	0.0	−1.414
8	0.0	1.414
9	0.0	0.0
10	0.0	0.0

$y(\eta_0, 45) = -125^2 g\eta_0^2 + 85 = 0 \Rightarrow \eta = +\sqrt{\frac{85}{125^2 g}}$, or in other words, he would need

$$v_0 = 125\sqrt{\frac{g}{85}} \approx 42.44\,\text{m/s}$$

If the launch angle, θ, is 45°, and initial velocity is 42.44 m/s, then the terminal time should equal:

$$t_f = \frac{125}{42.44\cos(45)} \approx 4.165\,\text{s}$$

Wow! This is great! Geoffrey was ecstatic. Now all he needed to do was figure out how many twists of the launching rope it would take to achieve this velocity. But, Geoff also realizes that he can alter the length of the throwing arm (lever), and that might affect the initial velocity. This problem calls for an experiment. Geoffrey has his intrepid crew construct a 120 m tower with a painted target centered at approximately 40 m off the ground. Then he had the crew move the catapult approximately 125 m away from the tower. Now, Geoff needs a plan. He decide to vary the number of twists (rotations of a tightening gear) between 10 and 15, and the length of the launch lever arm (which was cleverly designed to be continuously adjustable) between 2 and 3 m. Figure 7.6 shows a picture of Geoffrey's mangonal. Table 7.4 shows the various runs that Geoff and his crew decided to try.

Oddly enough, they chose a two-factor CCD, with two replicates at the center point. Note the axial points extend beyond the lower (−1) and upper levels (+1) for the two factors. Table 7.5 shows the same design in Helmert-coded units.

The axial points are located at $\pm(\sqrt{2})$ coded units away from the center point in each variable's direction.

Table 7.6 Catapult experimental runs with data

Rotations	Arm	$Y(t_f)$
10	2	35.3
10	3	77.6
15	2	84.2
15	3	103.8
9	2.5	49.0
16	2.5	101.2
12.5	1.8	57.0
12.5	3.2	98.4
12.5	2.5	81.5
12.5	2.5	85.2

So, if everything works according to plan, the optimal conditions would result in a flight time of approximately 4.165 s. As it turns out, time measurement was not quite up to the thousandths of a second in Geoff's day. In fact, Geoff could not even measure time to the nearest minute. What he could do was measure the height, in meters, at which the projectile struck the tower. This is precisely what he did. Of course, he wanted to reuse his practice tower, so instead of launching rocks, Geoff had his crew launch watermelons. They had the added advantage of leaving a nice mark where they hit, so the height could be easily measured. Table 7.6 shows the data he and his crew gathered.

Figure 7.7 shows the JMP output from fitting a second-order model to the data.

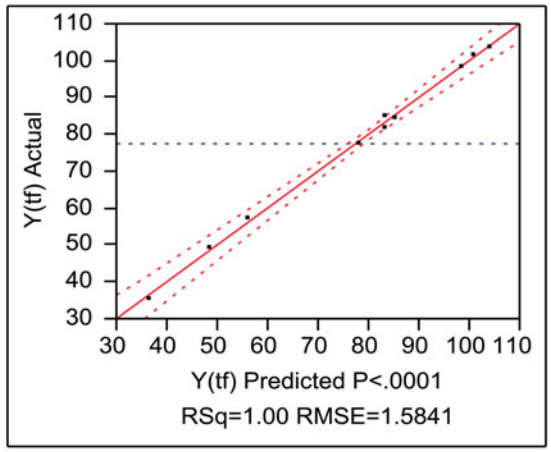
All the terms in the model had significant coefficients, and the adjusted R^2 was approximately 0.9953. Thus it looks like the model will provide reasonable approximations to the height at time of impact ($Y(t_f)$) response variable. Table 7.7 shows the actual data together with the model predictions.

All things considered, Geoff was pretty happy with his model. The next step was to find the optimal conditions for rotations and arm length, in order to hit the tower at approximately 40 m from the ground. Using JMP "Contour Profiler", we are able to see a contour in the approximating second-order "surface" where the height is equal to 40 m, approximately. Figure 7.8 shows the plot. The contour shown on the plot is the 40-m mark. Thus any point (Rotations, Arm length) on this contour is predicted to yield a striking height of 40 m. For example, if Rotations is set to about 9.62 turns, and Arm Length is approximately 2.17 m, then the model predicts the striking height from 125 m away would be at 40 m.

We have found that with approximately 9.62 rotations with an arm length of approximately 2.17 m will deliver the projectile at the 40 m mark (predicted value $\approx$ 40.19 m) on the keep of the castle, as long as the catapult is 125 m from the outer wall (and safely out of archery range). The 95 % confidence interval for this predicted value is approximately (37.00 m, 43.37 m). Since the window was 9 m in height, with the center at 40 m, they figured that they had a pretty good chance of getting the projectile to fly through it. However, to verify the optimal conditions, Geoffrey had his crew try out the optimal conditions. They attained a height of 40.48 m, which is a little high, but close enough.

Our EAS used factors in natural units to fit his approximating model. In this case, he was interested in obtaining a predictive equation, not in comparing magnitudes of coefficients. It was more intuitive to be able to relate the factors in their natural units to the response variable.

The reader might notice a few flaws in our analysis of the catapult problem. For one thing, the projectile is not exactly massless. Secondly, the projectile certainly will provide some air resistance. Thirdly, we have assumed that the window in the keep tower will be open when the projectile arrives. Well, these are all good points, but the bottom line is that we have employed a CCD to obtain an approximating equation for making design decisions.



Summary of Fit

RSquare	0.997917
RSquare Adj	0.995314
Root Mean Square Error	1.584079
Mean of Response	77.31769
Observations (or Sum Wgts)	10

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	4809.7117	961.942	383.3500
Error	4	10.0372	2.509	Prob > F
C. Total	9	4819.7489		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-418.4074	36.68011	-11.41	0.0003*
Rotations	36.730859	3.401491	10.80	0.0004*
Arm	150.09318	17.00746	8.83	0.0009*
Rotations*Arm	-4.556887	0.633632	-7.19	0.0020*
Rotations**2	-0.714163	0.120068	-5.95	0.0040*
Arm**2	-12.58032	3.00169	-4.19	0.0138*

Fig. 7.7 Second-order approximating model for catapult data

BBD

The Box-Behnken designs consist of "face center" points, together with a design center. A face center is a run in which one factor is set to its center value and the other factors are at one of their extreme (low or high) values. In a two-factor experiment (such as the catapult problem), there are four face centers. Table 7.8 shows the runs for a BBD.

Table 7.7 Catapult data with model predictions

Rotations	Arm	$Y(t_f)$	Model predictions
10	2	35.3	36.2
10	3	77.6	77.8
15	2	84.2	85.0
15	3	103.8	103.9
9	2.5	49.0	48.4
16	2.5	101.2	100.8
12.5	1.8	57.0	56.0
12.5	3.2	98.4	98.3
12.5	2.5	81.5	83.3
12.5	2.5	85.2	83.3

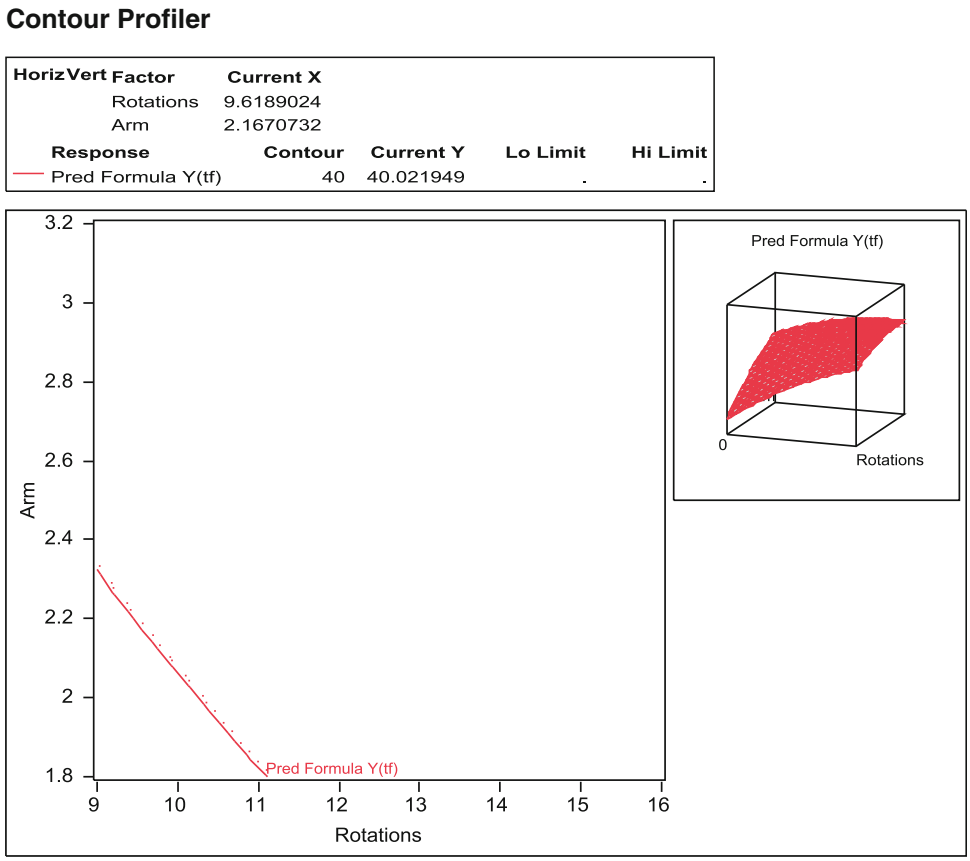


Fig. 7.8 Optimal contour—catapult model

A BBD will allow the analyst to fit a second-order model to the data, and it requires fewer design points than the CCD. As in the case of the CCD, it is common to replicate the design-center point. There are some disadvantages. Unlike the CCD, there is no two-level design "embedded" in the BBD (consider runs 1–4 in Tables 7.4 and 7.5). Should data be lost at the axial or center points of the CCD, at least a first-order model can be fit. The other potential issue is the fact that there are no axial points, per se, in the BBD. Thus, it might be helpful to extend the lower and upper limits of the experimental ranges for the factors in the BBD.

Table 7.8 BBD for a two-factor experiment

Run	Type	H1	H2
1	Face-center	−1	0
2	Face-center	+1	0
3	Face-center	0	−1
4	Face-center	0	+1
5	Design-center	0	0

Another Slight Digression: Hypothesis Tests About Model Parameters

Sometimes the EAS may want to test an hypothesis about a given parameter in a model. One question that the EAS might ask is the power provided by a particular design to test such an hypothesis. As an example, consider a second-order model.

Fit a model of the form:

$$y_{ijk} = \beta_0 + \sum_i \beta_i x_i + \sum_{i,j} \beta_{ij} x_i x_j + \sum_i \gamma_i x_i^2 + \varepsilon_{ijk}$$

Let **X** represent the design matrix, which has a column for each parameter in the model, and a row for each run and replicate. Usually, the levels of the regressors (x) are coded via Helmert coding. Since this is a second-order model, the levels will be coded to −1, 0, and +1 (assuming a designed experiment).

The least-squares estimates of the parameters will be called b_0, b_i, b_{ij}, and g_i, corresponding to the intercept, first order terms, cross-product terms, and the second order terms.

If **b** represents the vector of least-squares parameter estimates, then the variance-covariance matrix of the estimates is given by:

$$\mathbf{V}(\mathbf{b}) = \sigma^2 \left[\mathbf{X}'\mathbf{X}\right]^{-1}$$

where σ represents the standard deviation of noise, ε. This standard deviation is assumed to be constant over all observations. The standard error of any parameter component of the vector **b** is the square root of the corresponding diagonal component in the matrix **V**(**b**). Call the diagonal element of $[\mathbf{X}'\mathbf{X}]^{-1}$ c_{kk}^2, so that the standard error of b_k is:

$$SE(b_k) = \sigma c_{kk}$$

Under the null hypothesis that $\beta_k = \nu_0$, the statistic:

$$t = \frac{b_k - \nu_0}{\widehat{\sigma} c_{kk}}$$

has a Student's t distribution with degrees of freedom equaling the error degrees of freedom (*dfe*) from the model (assuming Gaussian noise, ε). The symbol $\widehat{\sigma}$ is the estimate of σ, which is usually the root mean square error from the model fitting. So, to test the hypothesis:

H_0: $\beta_k = \nu_0$ against the alternative H_1: $\beta_k \neq \nu_0$

Compare the statistic |t| to a 100(1−α/2) percentile of a central Student's t distribution, $t_{1-\alpha/2}$. If $|t| > t_{1-\alpha/2}$, then reject H_0.

Under the alternative hypothesis that $\beta_k = \lambda \neq \nu_0$, the statistic has a non-central t distribution (Johnson et al. 1995) with *dfe* degrees of freedom and non-centrality parameter:

Table 7.9 Three factor Box-Behnken design—from Minitab 16

StdOrder	RunOrder	PtType	Blocks	X1	*X2*	X3
1	1	2	1	−1	−1	0
2	2	2	1	1	−1	0
3	3	2	1	−1	1	0
4	4	2	1	1	1	0
5	5	2	1	−1	0	−1
6	6	2	1	1	0	−1
7	7	2	1	−1	0	1
8	8	2	1	1	0	1
9	9	2	1	0	−1	−1
10	10	2	1	0	1	−1
11	11	2	1	0	−1	1
12	12	2	1	0	1	1
13	13	0	1	0	0	0
14	14	0	1	0	0	0
15	15	0	1	0	0	0

Table 7.10 $X'X^{-1}$ for three factor Box Behnken design

	X0	X1	*X2*	X3	X12	X13	X23	X1S	X2S	X3S
X0	0.3333	0	0	0	0	0	0	−0.1667	−0.1667	−0.1667
X1	0	0.125	0	0	0	0	0	0	0	0
X2	0	0	0.125	0	0	0	0	0	0	0
X3	0	0	0	0.125	0	0	0	0	0	0
X12	0	0	0	0	0.25	0	0	0	0	0
X13	0	0	0	0	0	0.25	0	0	0	0
X23	0	0	0	0	0	0	0.25	0	0	0
X1S	−0.1667	0	0	0	0	0	0	0.2708	0.0208	0.0208
X2S	−0.1667	0	0	0	0	0	0	0.0208	0.2708	0.0208
X3S	−0.1667	0	0	0	0	0	0	0.0208	0.0208	0.2708

$$\frac{\lambda - \nu_0}{\widehat{\sigma} c_{kk}}$$

Note that the value of c_{kk} is dependent upon the number of runs and replicates. It is also dependent on the way in which the factors are coded.

Suppose the experiment consists of a three factor Box-Behnken design with Helmert coding, and replication only at the center point. Furthermore, suppose that three center points are run. Thus, the experiment consists of 15 runs, as illustrated in Table 7.9. The design matrix would include the columns X1, *X2*, and X3 in Table 7.9, and additional columns for the intercept, the three two-way cross-product terms and the three second-order terms. In total, the design matrix, **X**, would have 10 columns and 15 rows.

Table 7.10 shows the inverse of the matrix $\mathbf{X}'\mathbf{X}$, namely $\mathbf{X}'\mathbf{X}^{-1}$.

Note that **X0** represents the entry for the intercept, **X12**, **X13**, **X23** are the entries for the two-way cross-product terms, and **X1S**, **X2S**, and **X3S** are the entries for the squared terms.

We will consider testing the hypothesis H_0: $\beta_{23} = \nu_0 = 0$ against the alternative H_1: $\beta_{23} \neq 0$.

As an example, the value of $c_{77} = \sqrt{0.25}$ corresponds to the standard error of the coefficient for the X23 cross-product term (*X2**X3). The standard error for the estimate b_{23} is:

$$SE(b_{23}) = \sigma c_{77} = \sigma\sqrt{0.25}$$

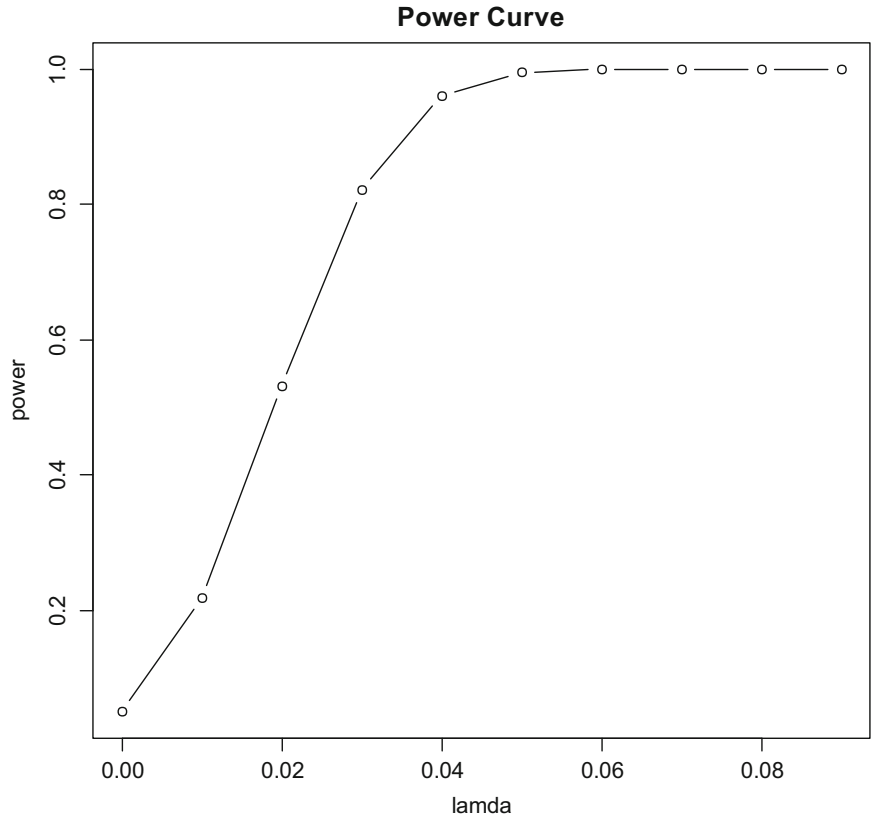


Fig. 7.9 Power curve for testing H_0: $\beta_{23} = 0$

In this design with a second-order model there are 15 – 10 = 5 degrees of freedom for error. If an a priori guess for the value of $\widehat{\sigma}$ can be obtained, then a power or operating characteristic (OC) curve (Grant and Leavenworth 1980), which is 1—power plotted against alternative values for the parameter, can be constructed for the hypothesis test.

The 97.5th percentile of a central Student's t distribution with five degrees of freedom is approximately 2.5706. The power is the probability that a non-central t statistic will exceed (in absolute value) this percentile:

$$Pr\left\{t'\left(df = 5, ncp = \frac{\lambda}{\widehat{\sigma}\sqrt{0.25}}\right) < -2.5706 \; OR \; t'\left(df = 5, ncp = \frac{\lambda}{\widehat{\sigma}\sqrt{0.25}}\right) > 2.5706\right\}$$
$$= Pr\left\{\left|t'\left(df = 5, ncp = \frac{\lambda}{\widehat{\sigma}\sqrt{0.25}}\right)\right| > 2.5706\right\}$$

Where t' represents a non-central t with df degrees of freedom and non-centrality parameter ncp. Note that ncp is a function of λ, the alternate value of the regression parameter. At $\lambda = \nu_0$, the probability is α. The power here is calculated for both positive and negative values of λ, but the x-axis only shows the absolute value of λ.

Figure 7.9 shows the power curve with an assumed $\widehat{\sigma} = 0.02$, and with the null hypothesis that H_0: $\beta_{23} = 0$.

Figure 7.10 shows the R script that was used to generate the power curve.

```
options(na.action=na.exclude)
#
# setwd tells R which folder to use for this script
#

setwd("H:\\Personal Data\\Scott\\Documents\\PROGRAMS\\R\\")

power <- c()
lamda <- c()
ncp <- c()

df1 <- read.csv("20141002 Box Behnken 3 Factor Design.csv")

xmat <- data.matrix(df1) #converts data frame into a matrix

sigmahat <- 0.02

nu0 <- 0.0

xmatprime <- t(xmat) #transpose function

xtx <- xmatprime %*% xmat

xtxinv <- solve(xtx)

se_b23 <- sqrt(xtxinv[7,7])*sigmahat

lamda <- c(nu0, nu0+0.01, nu0+0.02, nu0+0.03, nu0+0.04, nu0+0.05, nu0+0.06,
nu0+0.07, nu0+0.08, nu0+0.09)
delta <- lamda - nu0
crit_val <- qt(.95,5)
# Here we use the 95th percentile of the t-distribution
# so that we can plot the absolute value of delta, instead of negative and
# positive values.  That is, we only plot in the program non-negative values of delta
# with the understanding that we have sort of folded the power curve.
# the power curve ought to go from negative values of delta to positive values
#

ncp <- delta / se_b23 # this divides each component of delta by the scalar se_b23

power <- pt(crit_val, 5, ncp, lower.tail = FALSE)

plot(delta,power,type="b",main="Power Curve")
```

Fig. 7.10 R script for generating the power curve

```
# note that solve can also be used to solve systems
# if a left-hand side vector is also given as an input to the function
#
#you can plot power as a function of delta or lamda
#

# this writes xtxinv to a .csv file
#
write.csv(xtxinv,file="xtxinverse.csv")
```

Fig. 7.10 (continued)

Key Points

- In order to fit a second-order polynomial model, the factors/regressors must have at least three levels included in the experiment.
- A Center Point run with a two-level experiment can help detect the presence of higher order effects.
- A Central Composite Design (CCD) is a two-level factorial design with a center point and axial or face-centered points added.
- A Box-Behnken Design (BBD) requires fewer points than a CCD, but does not include the "corner points" of a two-level design.
- Both CCDs and BBDs are rotatable.
- Confidence intervals and hypothesis tests can be computed for model parameters.

Exercises and Questions

1. How many unique runs (not including any replicates) would a Box-Behnken design in five factors require?
2. Would you ever consider using a Central Composite Design with 10 factors? If yes, in what situation?

Chapter 8
Mixture Experiments

The basic mixture/formulation problem is to decide, for a given set of components, how much of each component should be put into the mixture. Commonly, the quantity of the mixture is fixed (either by weight or volume) so that the amount of each component can be expressed as a fraction or percent of the whole mixture.

Generally there is a response variable that is an unknown function of the distribution of components, also referred to as mixture variables. The EAS must use empirical data to find an approximating equation to describe the relationship between the response and the mixture variables. So far, this is sounding a lot like the problems in factorial experimentation. The big difference is that we cannot choose the "levels" of the mixture variables independently of each other. That is, if there are only two components, and one comprises 25 % of the mixture, then the other must comprise 75 %. So, in experimentation with mixture variables, we cannot simply choose low and high values, and pick various combinations of low and high levels for each of the mixture variables.

Mixture variables, call them x_k, satisfy the following constraints:

$$x_k \geq 0, k = 1, q,$$

$$\sum_{i=1}^{q} x_i = 1.$$

There are several points to mention about these constraints. First, it is allowable for any given mixture variable (also called components) to equal 0 or 1. Secondly, a mixture is defined by a combination of values for the mixture variables. Thirdly, the constraint that all the values of the components add to 1 could be slightly altered, to say that they must add to a fixed value, say A. In other words, the summation constraint could be:

$$\sum_{i=1}^{q} x_i = A.$$

For our discussions, we will assume that the mixture variables have been "normalized", so that their values are proportions of the total mixture.

The constraints on mixture variables creates a geometric structure for the space in which mixtures are defined. The structure is called a simplex. While the topological and geometric properties of the simplex are fascinating, we will not discuss them, except to elucidate properties of experiments. As in

S.A. Pardo, *Empirical Modeling and Data Analysis for Engineers and Applied Scientists*,
DOI 10.1007/978-3-319-32768-6_8

the case of factorial experiments, we will not provide a complete exposition of mixture experiments and their associated analyses. We will cover some basic ideas.

The First-Order Model

Suppose the EAS must create a mixture of q components. Suppose further that the quality of any mixture can be assessed by measuring a response variable, Y. Perhaps the EAS wants to maximize Y, or minimize it, or even have Y fall between two specification limits. The first step is to be able to predict the value of Y given a particular mixture. So, as in the case of the factorial experiments, the EAS needs an approximating equation, or model. A first-order model might look like:

$$y_j = \beta_0 + \sum_{i=1}^{q} \beta_i x_i + \epsilon_j \text{ for } j = 1, m.$$

This implies that in order to estimate the $q + 1$ coefficients, $m \geq q + 1$. Now consider the equivalent model:

$y_j = \beta_0 \sum_{i=1}^{q} x_i + \sum_{i=1}^{q} \beta_i x_i + \epsilon_j$, which is equivalent since $\sum_{i=1}^{q} x_i = 1$. Rearranging terms gives: $y_j = \sum_{i=1}^{q} (\beta_0 + \beta_i) x_i + \epsilon_j$ and making the assignment $\alpha_i = \beta_0 + \beta_i$ we can express the first-order model as:

$y_j = \sum_{i=1}^{q} \alpha_i x_i + \epsilon_j$ which has only q parameters to estimate. The interpretation of this "reduced" model is also more intuitive, perhaps. After all, the response variable must equal 0 if all the components are set to 0 (i.e., there is no mixture on which to make any measurements).

Example: First-Order Model

Suppose a mixture may be made of exactly two components, x_1 and x_2. Furthermore, suppose it is possible to use only one of either of these two components as the entire mixture. That is, suppose the mixture could consist of 100 % of component one ($x_1 = 1, x_2 = 0$) or vice versa ($x_1 = 0, x_2 = 1$). The classic example is mixing different octane-numbered gasolines, and measuring the miles per gallon (we will use the English Standard units for this example, rather than metric or SI units). So the experimental design will be mixtures of two types of gasoline, high octane and regular octane. Let x_1 represent the proportion of the tank filled with high octane gas, and x_2 the proportion of the tank filled with regular. The response, Y will represent the miles per gallon, computed as the number of miles driven divided by gallons used. The next question is how to collect the data that will allow us to generate the approximating equation.

Here is one thought. Suppose we try one case where only the high octane gas is used ($x_1 = 1$, $x_2 = 0$) and another with only regular octane gas ($x_1 = 0, x_2 = 1$). Suppose further we replicate the experiment. Table 8.1 shows the data.

Figure 8.1 shows SAS code and output for fitting the first-order model.

The fitted model is:

$$y = 30.54x_1 + 24.17x_2$$

Since $x_1 + x_2 = 1$ (one of the mixture constraints), the model clearly indicates that the optimal mixture (highest mpg) is predicted to be $x_1 = 1, x_2 = 0$.

Table 8.1 Octane experiment—2-Point design

High octane x1	Regular octane x2	mpg y
1	0	30.9
0	1	22.0
1	0	30.2
0	1	26.4

One thing that may be bothering you is that our model was a linear function fit to two points, namely ($x_1 = 1$, $x_2 = 0$) and ($x_1 = 0$, $x_2 = 1$). We had no observations of mpg for any non-pure mixtures. Suppose we add the point, $x_1 = 0.5$, $x_2 = 0.5$. The data are shown in Table 8.2. They are graphically depicted in Fig. 8.2. The SAS output for the refitted first-order model is given in Fig. 8.3.

Notice that the mpg results at the 50/50 mixture was higher than any of the other observations.

There is not a big change. The fitted model is now:

$$y = 33.57x_1 + 27.20x_2$$

It appears that the optimal mixture is still 100 % high octane ($x_1 = 1$, $x_2 = 0$). Yet the data seem to indicate that this may not be the case. In factorial experiments, the midpoint, or centroid, data may indicate that a first-order model is inadequate.

The Second-Order Model

In the case with independent factors, a second-order model might look like:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \epsilon$$

We have already discussed eliminating the intercept in the case of mixture variables. Since the values of the mixture variables are dependent on each other (since their total must add to 1), notice that:

$$x_1^2 = x_1(1 - x_2) = x_1 - x_1 x_2 \text{ and } x_2^2 = x_2(1 - x_1) = x_2 - x_1 x_2$$

Replacing these expressions in the model (assuming we have already eliminated the intercept):

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_3 x_1 + \beta_3 x_1 x_2 + \beta_4 x_2 - \beta_4 x_1 x_2 + \epsilon$$

Collecting terms:

$$y = (\beta_1 + \beta_3) x_1 + (\beta_2 + \beta_4) x_2 + (\beta_{12} - \beta_3 - \beta_4) x_1 x_2 + \epsilon$$

Relabeling the parameters leaves us with:

$$y = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_{12} x_1 x_2 + \epsilon$$

The second-order model went from having six parameters to having only three parameters to estimate. Our old friend William Ockham would be proud. Of course, it is half as expensive to estimate three parameters compared to estimating six. The two-variable cross-product terms are referred to as "quadratic" terms in the context of mixture experiments.

```
libname stuff 'H:\Personal Data\Experimentation for Design &
Validation\Data & Analyses';

data calc;

  set stuff.d20131124_first_order_octane_exp; /* do not include the
sas7bdat extension here */
/*************************************************************
*                                                            *
* variables: x1 = high octane                                *
*            x2 = regular                                    *
*            y =  mpg                                        *
*                                                            *
*                                                            *
*                                                            *
*************************************************************/

  run;

proc sort data=calc;
  by x1 x2;
  run;

proc means data=calc;
  var y;
  by x1 x2;
  run;


proc glm data=calc;

  model y = x1 x2/est noint;

  run;
```

```
                              The SAS System         07:04 Sunday, November 24, 2013   4

------------------------------------------ x1=0 x2=1 ------------------------------------------

                                The MEANS Procedure

                              Analysis Variable : y y

          N            Mean         Std Dev         Minimum         Maximum
          ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff
          2      24.1745710       3.0843234      21.9936250      26.3555170
          ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff

------------------------------------------ x1=1 x2=0 ------------------------------------------

                              Analysis Variable : y y

          N            Mean         Std Dev         Minimum         Maximum
          ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff
          2      30.5417982       0.5130367      30.1790264      30.9045699
          ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff
```

Fig. 8.1 Octane experiment—SAS code and output

```
                                The SAS System         07:04 Sunday, November 24, 2013   5

                              The GLM Procedure

                    Number of Observations Read           4
                    Number of Observations Used           4

                                The SAS System         07:04 Sunday, November 24, 2013   6

                              The GLM Procedure

                       General Form of Estimable Functions

                           Effect        Coefficients

                           x1            L1

                           x2            L2

                                The SAS System         07:04 Sunday, November 24, 2013   7

                              The GLM Procedure

Dependent Variable: y   y

                                              Sum of
          Source                     DF       Squares     Mean Square    F Value    Pr > F

          Model                       2    3034.422639    1517.211319     310.39    0.0032

          Error                       2       9.776257       4.888129

          Uncorrected Total           4    3044.198896

                            R-Square     Coeff Var      Root MSE        y Mean

                            0.996789      8.081352      2.210911      27.35818

NOTE: No intercept term is used: R-square is not corrected for the mean.

          Source                     DF      Type I SS     Mean Square    F Value    Pr > F

          x1                          1    1865.602872    1865.602872     381.66    0.0026
          x2                          1    1168.819766    1168.819766     239.11    0.0042

          Source                     DF     Type III SS     Mean Square    F Value    Pr > F

          x1                          1    1865.602872    1865.602872     381.66    0.0026
          x2                          1    1168.819766    1168.819766     239.11    0.0042

                                                    Standard
                    Parameter         Estimate          Error    t Value    Pr > |t|

                    x1            30.54179818     1.56335036      19.54      0.0026
                    x2            24.17457100     1.56335036      15.46      0.0042
```

Fig. 8.1 (continued)

Fitting the second-order model to the data in Table 8.2 is accomplished using the SAS Code and its associated output in Fig. 8.4.

So now the predictive equation is:

$$y = 30.54x_1 + 24.17x_1 + 36.37x_1x_2$$

Table 8.2 Octane experiment—3-Point design

High octane x1	Regular octane x2	mpg y
1	0	30.9
0	1	22.0
1	0	30.2
0	1	26.4
0.5	0.5	37.9
0.5	0.5	35.0

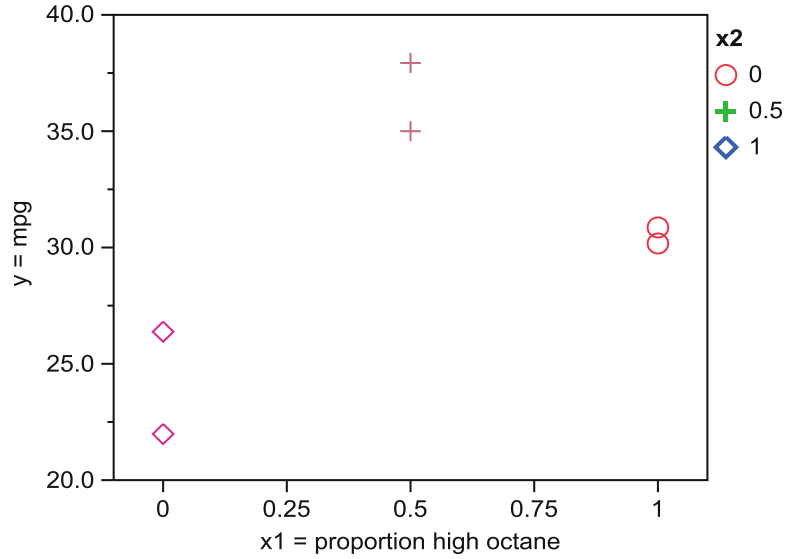


Fig. 8.2 Graphic presentation of octane experiment data

Table 8.3 shows the data together with predicted values from both first-order and second-order models.

It appears that the second-order model gives better predictions, at least for those conditions used to fit the model.

Another question we might ask is, what is the optimal mixture? Clearly this optimization must be performed subject to the mixture constraints, so the problem could be described as a nonlinear program:

Maximize: $y = 30.54x_1 + 24.17x_1 + 36.37x_1x_2$

Subject to:

$$x_1 \geq 0$$
$$x_2 \geq 0$$
$$x_1 + x_2 = 1$$

There are many solver programs. Figure 8.5 shows a screen shot from JMP Profiler which can be used to find optimal values for factors or mixture variables. Notice that the constraint:

$$x_1 + x_2 = 1$$

was added. The non-negativity constraints did not need to be added, since the JMP profiler will only plot the objective function (the model) over the range of values observed for the mixture variables.

```
The SAS System          07:04 Sunday, November 24, 2013  14

                                The GLM Procedure

                        General Form of Estimable Functions

                            Effect          Coefficients

                            x1              L1

                            x2              L2


                                  The SAS System          07:04 Sunday, November 24, 2013  15

                                The GLM Procedure

Dependent Variable: y   y

                                                 Sum of
Source                          DF             Squares      Mean Square    F Value    Pr > F

Model                            2          5581.412829      2790.706414      89.88    0.0005

Error                            4           124.196067        31.049017

Uncorrected Total                6          5705.608896


                      R-Square     Coeff Var      Root MSE        y Mean

                      0.978233      18.33625      5.572164      30.38879

NOTE: No intercept term is used: R-square is not corrected for the mean.

Source                          DF        Type I SS     Mean Square    F Value    Pr > F

x1                               1       3805.120968     3805.120968     122.55    0.0004
x2                               1       1776.291861     1776.291861      57.21    0.0016

Source                          DF       Type III SS     Mean Square    F Value    Pr > F

x1                               1       2705.055035     2705.055035      87.12    0.0007
x2                               1       1776.291861     1776.291861      57.21    0.0016

                                                    Standard
            Parameter                Estimate           Error    t Value    Pr > |t|

            x1                   33.57240332      3.59681670       9.33      0.0007
            x2                   27.20517614      3.59681670       7.56      0.0016
```

Fig. 8.3 Octane experiment—SAS output with centroid point added

Figure 8.6 shows the profiler graphs in greater detail. Note that the optimal mixture is $x_1 = 0.58$ (58 % high octane) and $x_2 = 0.42$ (42 % regular octane). The expected maximum response is 36.73 mpg.

Constraints in Mixture Designs

Typically mixtures or formulations have constraints on the components. Rarely are there mixture problems that allow for "pure blends", i.e., a mixture where $x_i = 1$ and $x_j = 0, j \neq i$. In fact, there can be many different constraints. Perhaps most simply are those constraints where some or all

```
libname stuff 'H:\Personal Data\Experimentation for Design &
Validation\Data & Analyses';
data calc;
  set stuff.d20131124_first_order_octane_mid; /* do not include the
sas7bdat extension here */
/*************************************************************
*                                                            *
* variables: x1 = high octane                                *
*            x2 = regular                                    *
*            y =  mpg                                        *
*                                                            *
*                                                            *
*                                                            *
*************************************************************/
  run;
proc glm data=calc;
  model y = x1 x2 x1*x2/est noint;
  run;
```

```
The GLM Procedure

                          Number of Observations Read           6
                          Number of Observations Used           6

                                The SAS System        07:48 Tuesday, November 26, 2013   3

                             The GLM Procedure

                       General Form of Estimable Functions

                           Effect        Coefficients

                           x1            L1

                           x2            L2

                           x1*x2         L3
```

Fig. 8.4 Octane experiment—second-order model

The SAS System 07:48 Tuesday, November 26, 2013 4

The GLM Procedure

Dependent Variable: y y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	5691.627639	1897.209213	407.09	0.0002
Error	3	13.981257	4.660419		
Uncorrected Total	6	5705.608896			

R-Square	Coeff Var	Root MSE	y Mean
0.997550	7.103937	2.158800	30.38879

NOTE: No intercept term is used: R-square is not corrected for the mean.

Source	DF	Type I SS	Mean Square	F Value	Pr > F
x1	1	3805.120968	3805.120968	816.48	<.0001
x2	1	1776.291861	1776.291861	381.14	0.0003
x1*x2	1	110.214810	110.214810	23.65	0.0166

Source	DF	Type III SS	Mean Square	F Value	Pr > F
x1	1	1865.602872	1865.602872	400.31	0.0003
x2	1	1168.819766	1168.819766	250.80	0.0005
x1*x2	1	110.214810	110.214810	23.65	0.0166

Parameter	Estimate	Standard Error	t Value	Pr > |t|
x1	30.54179818	1.52650240	20.01	0.0003
x2	24.17457100	1.52650240	15.84	0.0005
x1*x2	36.36726163	7.47830393	4.86	0.0166

Fig. 8.4 (continued)

Table 8.3 Octane experiment data with predictions

High octane x1	Regular octane x2	mpg y	First order	Second order
1	0	30.9	33.6	30.5
0	1	22.0	27.2	24.2
1	0	30.2	33.6	30.5
0	1	26.4	27.2	24.2
0.5	0.5	37.9	30.4	36.4
0.5	0.5	35.0	30.4	36.4

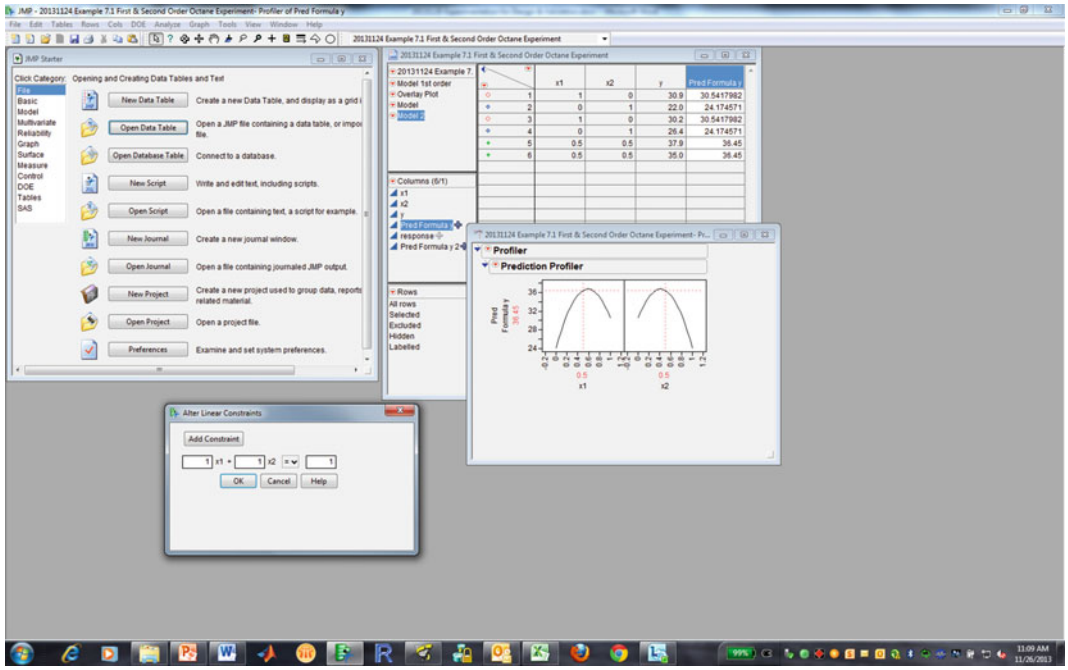


Fig. 8.5 Optimization of second-order model

components must comprise some minimum proportion of the mixture and no more than some maximum:

$$L_i \leq x_i \leq U_i, i = 1, q$$

Choosing the particular combinations of mixture variable values to include in an experiment becomes more complicated. For one thing, the summation constraint:

$$\sum_{i=1}^{q} x_i = 1$$

must still be satisfied for each mixture included in the experiment. Therefore, constraints placed on individual mixture variables must be feasible, in the sense that the summation constraint must be satisfied, as well as all the other constraints. For example, suppose we had decided to constrain the high octane variable to be no less than 60 % of the mixture, and at the same time we had constrained the regular octane variable to be no less than 50 % of the mixture. These two constraints would be infeasible, also referred to as inconsistent. When there are only two mixture variables, determining

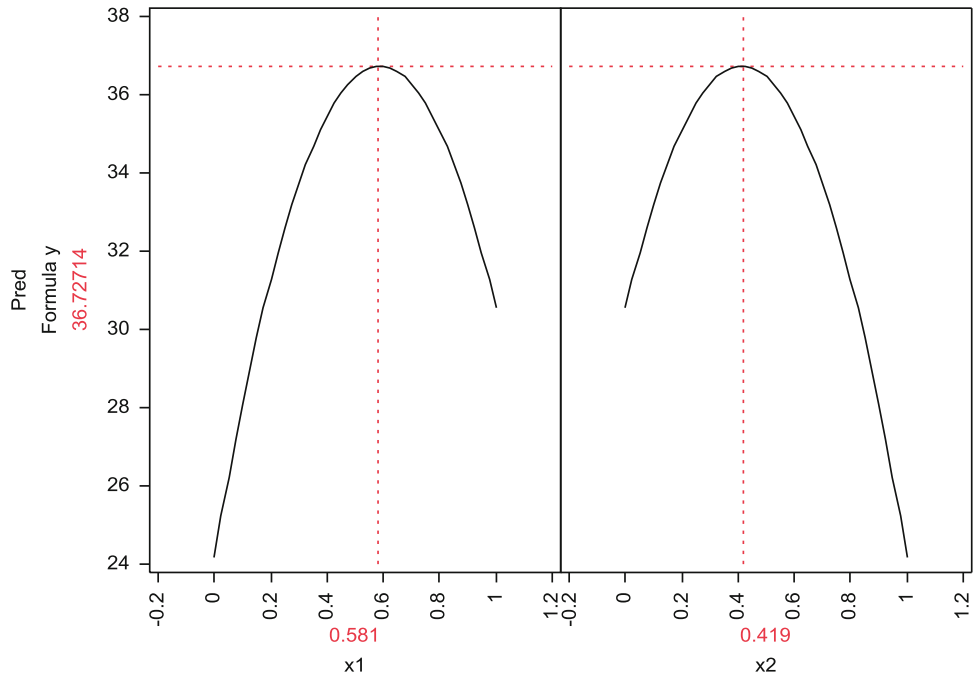


Fig. 8.6 Optimal mixture from second-order model

whether or not constraints may be inconsistent is fairly simple. For mixtures with more than two components, where constraints are expressed as lower and upper bounds, a procedure can be followed to detect whether constraints are inconsistent.

1. Compute the ranges, $R_i = U_i - L_i$, $i = 1, q$
2. If:

$$R_i > 1 - \sum_{j=1}^{q} L_j \text{ or } L_i + \sum_{j \neq i} U_j < 1.$$

Then the constraint U_i is said to be unattainable.

3. If:

$$R_i > \sum_{i=1}^{q} U_i - 1.$$

Then the constraint L_i is said to be unattainable.

If any of the constraints are unattainable, then the constraints are inconsistent. The values must be adjusted by raising upper limits or lowering lower limits.

Optimal Design

Generally, computers are used to evaluate algorithms for finding sets of mixtures that satisfy constraints, and adjusting constraints to make them consistent. The experimental designs are usually chosen to satisfy an optimality criterion based on standard errors, either standard errors of the

coefficient estimates, or standard errors of predicted values. The most common, and easiest to satisfy, is called D-optimality. A D-optimal design is a set of points chosen to minimize the volume of the joint confidence region of the model parameters. A G-optimal design is a set of points that minimizes the maximum standard error of a predicted value (within the set of points in the design). A V-optimal design minimizes the average standard error of predicted values (again, within the design points). In JMP, V-optimality is called I-optimality.

Optimal designs are optimal with respect to the order of the model being fit. Let X represents a matrix of design points (i.e., a set of mixtures), where the columns are assigned to each of the terms in the model, and the rows are the values each term takes for a given mixture. Then these optimality criteria are realized by doing the following:

D-Optimality: choose a set of mixtures that minimizes the determinant of the matrix $\left[\boldsymbol{X}'\boldsymbol{X}\right]^{-1}$. This minimizes the imprecision of estimation for the model parameters.

G-Optimality: choose a set of mixtures that minimizes the maximum value of $\boldsymbol{x}'\left[\boldsymbol{X}'\boldsymbol{X}\right]^{-1}\boldsymbol{x}$. This minimizes the width of confidence intervals for an individual predicted value.

V-Optimality: choose a set of points that minimizes $\frac{1}{\boldsymbol{n}}\boldsymbol{x}'\left[\boldsymbol{X}'\boldsymbol{X}\right]^{-1}\boldsymbol{x}$, where n = the number of mixtures included in the design. This minimizes the width of confidence intervals for the average predicted value.

In practice, the optimal designs are chosen by computer program for a fixed number of points, or mixtures, to include in the experiment. So, before an optimal design may be chosen, the experimenter must specify to the software both a model order and a number of design points.

A particularly excellent text on mixture experiments is that of Wendell F. Smith (2005).

Key Points

- Mixture experiments involve factors whose "levels" add to a constant that is the same for each run; usually the constant is 1.
- Second-order mixture experiments have second-order cross-product terms, but no squared terms.
- Analyses for mixture experiments are the same as those for designs with factors whose levels are independently chosen from one another.
- The choice of design can be done to satisfy various optimality criteria, such as minimizing the parameter estimate variance (D-optimal), minimizing individual predicted value variance (G-optimal), or minimizes average predicted value variance (V-optimal).

Exercises and Questions

1. Create a mixture design for fitting a second-order model with four mixture components, X1, X2, X3, X4, and the following constraints:

$$\begin{aligned} &\mathrm{X1} \geq 0.025 \\ &\mathrm{X2} + \mathrm{X3} \leq 0.60 \\ &\mathrm{X4} \geq 0.15 \end{aligned}$$

2. What sort of designs might be constructing if there are both mixture components and processing variables?

Chapter 9
Some Examples and Applications

Range Finding

Once an optimal point has been found, it may be desirable to find ranges of the factors where the response may not be optimal but may be sufficient. The approach we will take is based on the notion of inverse regression (Draper & Smith, 1998).

Suppose a first-order model has been fit:

$$\widehat{y} = b_0 + b_1 x$$

Suppose further that we have determined a range of values for the response variable, y, that are acceptable. Let the range be described as:

$$y_0 \pm \delta$$

The range of x values expected to yield acceptable values of the response are:

$$x_L = \frac{y_0 - \delta - b_0}{b_1}$$

and

$$x_U = \frac{y_0 + \delta - b_0}{b_1}$$

The estimation issue is that since both b_0 and b_1 have sampling variation, we would like to obtain some kind of probabilistic bounds on x_L and x_U. Using the standard error formula for a predicted value from a simple linear regression, the lower confidence limit for a "predicted" value of $y_0 - \delta$ would be given by:

$$\widehat{y}_L = y_0 - \delta - ts\sqrt{\frac{1}{n} + \frac{(x_{LL} - \overline{x})^2}{SS_x}}$$

where $SS_x = \sum_{i=1}^{n} (x_i - \overline{x})^2$, t is the appropriate percentile of a Student's t distribution, $\widehat{y}_L = b_0 + b_1 x_L$, and s is the RMSE from the regression. Solving for x_{LL} entails finding the roots of

S.A. Pardo, *Empirical Modeling and Data Analysis for Engineers and Applied Scientists*,
DOI 10.1007/978-3-319-32768-6_9

a quadratic polynomial, and using the smaller real root as the lower confidence limit on x_L. A similar process could be followed for x_{UU}, namely, solving:

$$\widehat{Y}_U = y_0 + \delta + ts\sqrt{\frac{1}{n} + \frac{(x_{UU} - \overline{x})^2}{SS_x}}$$

For the sake of being conservative, use the larger real root as the upper confidence limit for x_U. In the cases of no real roots, one could use $\pm|z|$, where z is the complex root. The idea of using the confidence limits for the potential range of x values is to provide the experimenter with a set of experimental conditions to use in an exploratory fashion. That is, once the range of x values is determined algebraically, an experiment should be performed, to verify that the response values are in fact acceptable at the extended limits on x.

More generally, suppose the polynomial model is fit:

$$\widehat{y} = b_0 + b_1 Z_1 + b_2 Z_2 + \ldots + b_k Z_k$$

Note that the Z_j could be powers of a single regressor, or they could be other variables. The inverse regression problem would be to find values of the Z_j, call them Z_j^0, such that

$$\widehat{y}_0 = b_0 + b_1 Z_1^0 + b_2 Z_2^0 + \ldots + b_k Z_k^0$$

To obtain point estimates for the ranges of the regressors, solve:

$$\widehat{y}_0 - \delta = b_0 + b_1 Z_1^L + b_2 Z_2^L + \ldots + b_k Z_k^L$$

and

$$\widehat{y}_0 + \delta = b_0 + b_1 Z_1^U + b_2 Z_2^U + \ldots + b_k Z_k^U$$

Finding confidence limits for the Z_j^L and the Z_j^U would involve finding solutions to:

$$t^2 SE^2 = \left(b_0 + b_1 Z_1^{LL} + b_2 Z_2^{LL} + \ldots + b_k Z_k^{LL} - (\widehat{y}_0 - \delta)\right)^2$$

and

$$t^2 SE^2 = \left(b_0 + b_1 Z_1^{UU} + b_2 Z_2^{UU} + \ldots + b_k Z_k^{UU} - (\widehat{y}_0 + \delta)\right)^2$$

$$SE = s\sqrt{z_0'\left[\boldsymbol{Z}'\boldsymbol{Z}\right]^{-1} z_0}$$

$$z_0 = \begin{bmatrix} 1 \\ Z_1^0 \\ \vdots \\ Z_k^0 \end{bmatrix}$$

$$\boldsymbol{Z} = \text{matrix of regressor values}$$
$$s = \text{RMSE from regression}$$
$$t = \text{appropriate percentile of Student's t distribution.}$$

A Quadratic Example

Suppose the equation resulted from a least squares fit:

$$\hat{y} = 10.39 + 38.10x - 48.80x^2$$

Figure 9.1 illustrates the data and the approximating polynomial.

The root mean square error (RMSE) from the model was $s = 0.7124$. If x_0 represents the optimal point, then the estimate of the optimal response is the solution to:

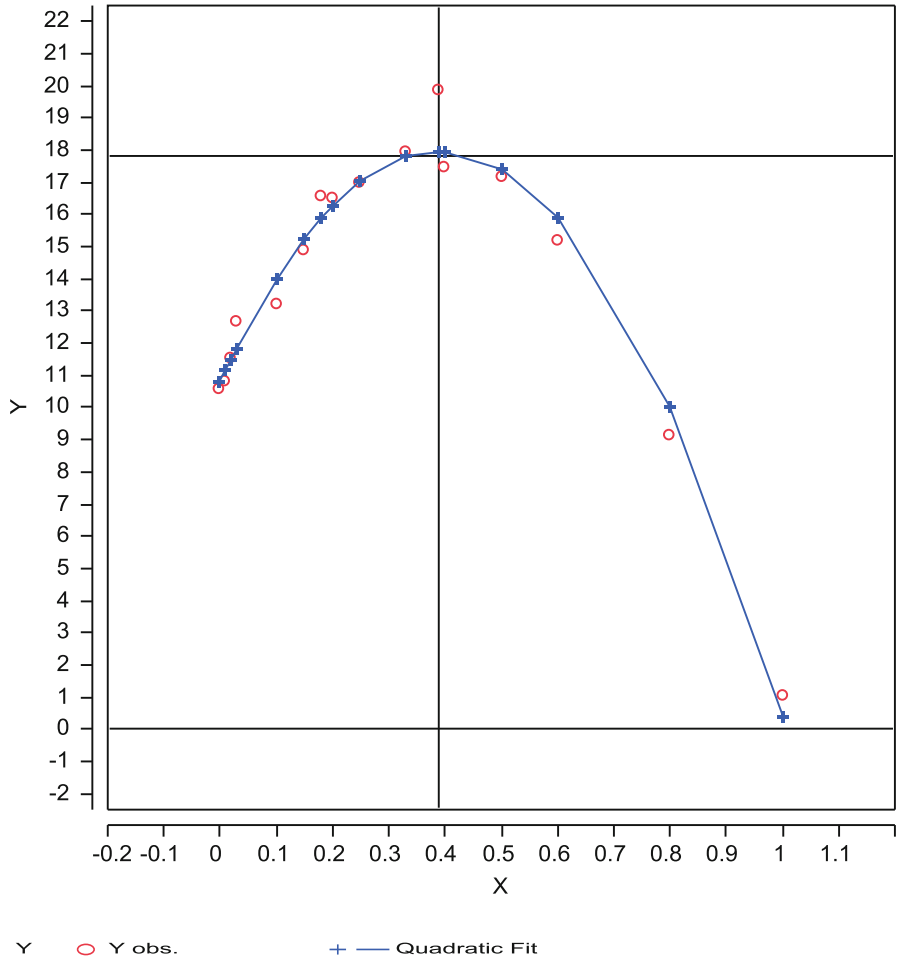


Fig. 9.1 Data and polynomial fit for quadratic range finding example

$$38.10 - 48.80x_0 = 0, x_0 \approx 0.3904, \ \widehat{y}_0 \approx 17.83$$

Suppose that any value of the response within $\widehat{y}_0 \pm \delta = 17.83 \pm 0.15$ would be acceptable. Then all that is necessary to find values of x such that:

$$b_0 + b_1 x + bx^2 = y_0 - \delta \tag{9.1}$$

or

$$b_0 + b_1 x + bx^2 = y_0 + \delta \tag{9.2}$$

Since the value of y_0 is a maximum for this quadratic function, the solutions to equation (9.1) are what we seek. The solutions are therefore $x_L \approx 0.3349$ and $x_U \approx 0.4458$. That is, these are "point estimates" for the limits of acceptable x values. However, a 100(1-α)% confidence limit for $y_0 - \delta$ would be:

$\widehat{y}_0 - \delta - tSE\left(\widehat{y}_0 | x_0\right)$ where t is the appropriate percentile of a Student's t distribution and

$$SE\left(\widehat{y}_0 | \boldsymbol{x}_0\right) = s\sqrt{\boldsymbol{x}_0'\left[\boldsymbol{X}'\boldsymbol{X}\right]^{-1}\boldsymbol{x}_0}$$

and

$$\boldsymbol{x}_0 = \begin{bmatrix} 1 \\ x_0 \\ x_0^2 \end{bmatrix}$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_{12} & x_{12}^2 \end{bmatrix}$$

As usual, s represents the RMSE from the regression. To find confidence limits x_L or x_U, find the solutions to the equation:

$$b_0 + b_1 x + bx^2 = y_0 - \delta - tSE\left(\widehat{y}_0 | x_0\right)$$

With

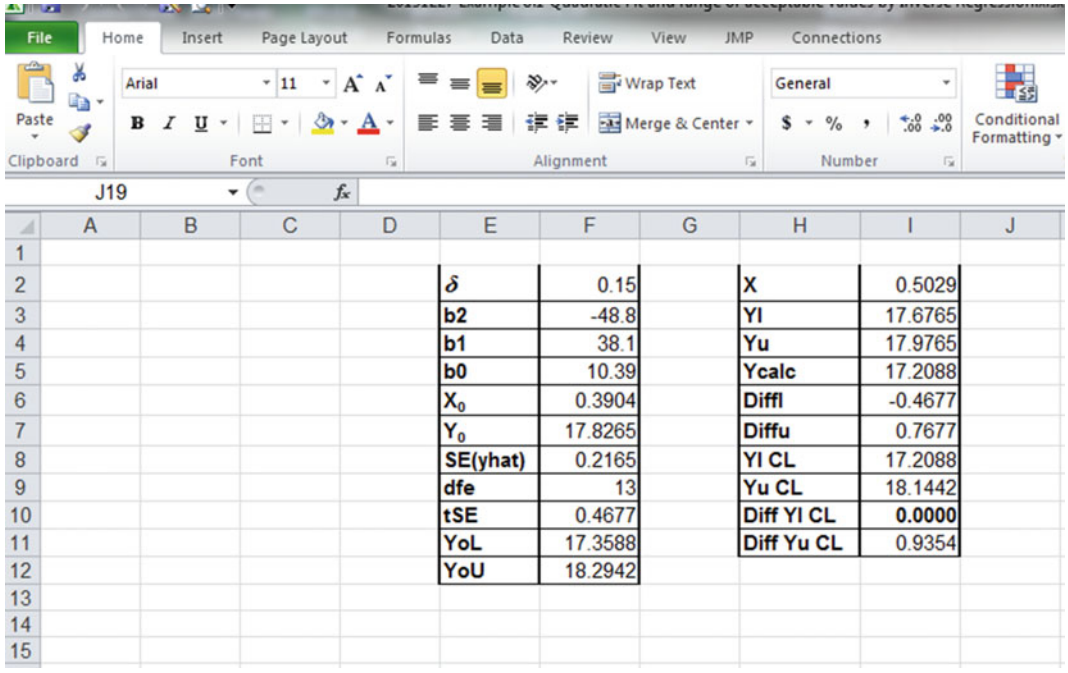
$$SE\left(\widehat{y}_0 | \boldsymbol{x}_0\right) \approx 0.2165$$

then:

$$\widehat{y}_0 - \delta - tSE\left(\widehat{y}_0 | x_0\right) \approx 17.2088.$$

The 95 % confidence limits are $x_L \approx 0.2779$ and $x_U \approx 0.5029$. This interval may be interpreted as the range of x values which is 95 % certain of yielding an acceptable response value.

Once the range of x values has been determined, a confirmatory experiment is appropriate. That is, observe the value of y at x_L and at x_U and check to see if your observations actually are within the desired range of the response.



	A	B	C	D	E	F	G	H	I	J
1										
2					δ	0.15		X	0.5029	
3					b2	-48.8		Yl	17.6765	
4					b1	38.1		Yu	17.9765	
5					b0	10.39		Ycalc	17.2088	
6					X_0	0.3904		Diffl	-0.4677	
7					Y_0	17.8265		Diffu	0.7677	
8					SE(yhat)	0.2165		Yl CL	17.2088	
9					dfe	13		Yu CL	18.1442	
10					tSE	0.4677		Diff Yl CL	0.0000	
11					YoL	17.3588		Diff Yu CL	0.9354	
12					YoU	18.2942				
13										
14										
15										

Fig. 9.2 Excel set-up for finding limits with quadratic response equation

By accounting for sampling variation in the predicted value, the solutions x_L and x_U may yield response values that are outside the acceptable range of performance. However, by performing an additional experiment over the range (x_L, x_U), the EAS may find that there is a wider interval in x, beyond the "point estimate" limits, over which acceptable response values can be obtained with a high degree of probability.

In the case of the quadratic approximating equation in a single factor, solving for the limits (x_L, x_U) can be accomplished with the quadratic solution formula. When there are more than one regressors, or there are higher-order terms, such formulas may not exist. Numerical methods can be used, however. The MS Excel™ Solver Add-in provides a tool for solving maximization/minimization/root-finding problems subject to constraints. Figure 9.2 shows an Excel worksheet with a set-up for finding limits using Solver.

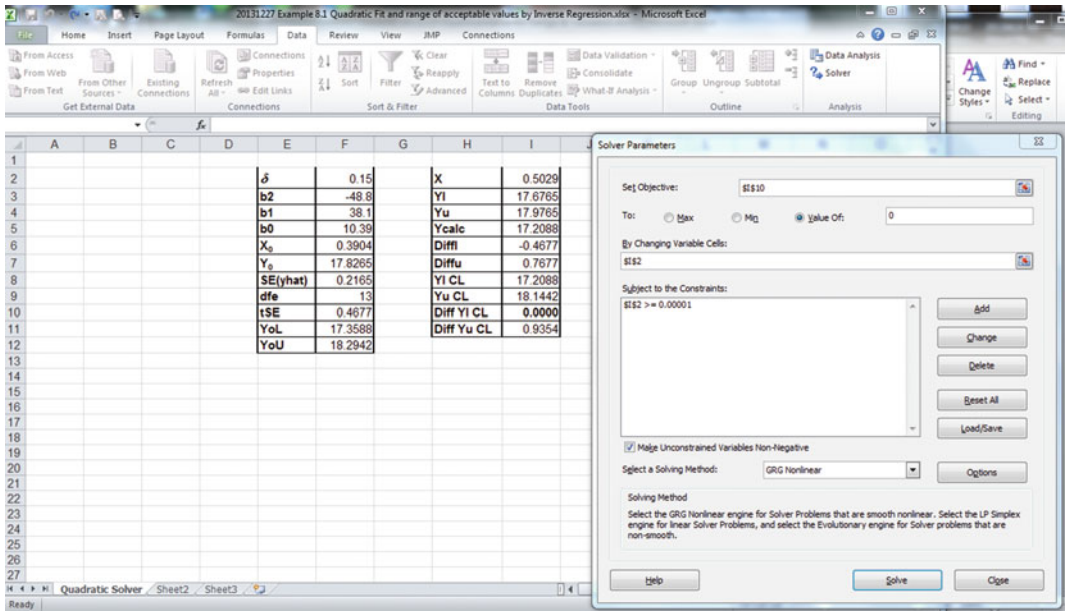
Figure 9.3 shows the Solver window.

Since there are two solutions to the equation, the Solver must be executed twice, with a starting value for x (**X**) that is lower than desirable, and a starting value of x that is higher than desirable. In this case, the objective is to set the function:

$$\mathbf{Diff\ Yl\ CL} = \mathbf{Ycalc} - (\mathbf{Y}_0 - \mathbf{tSE} - \delta)$$

to equal 0, by changing the value of **X. Ycalc** is given by:

$$\mathbf{Ycalc} = \mathbf{b}0 + \mathbf{b}1{*}\mathbf{X} + \mathbf{b}2{*}\mathbf{X}^2$$



δ	0.15
b2	-48.8
b1	38.1
b0	10.39
X_0	0.3904
Y_0	17.8265
SE(yhat)	0.2165
dfe	13
tSE	0.4677
YoL	17.3588
YoU	18.2942

X	0.5029
Yl	17.6765
Yu	17.9765
Ycalc	17.2088
Diffl	-0.4677
Diffu	0.7677
Yl CL	17.2088
Yu CL	18.1442
Diff Yl CL	0.0000
Diff Yu CL	0.9354

Fig. 9.3 Solver screen

Table 9.1 Blender experimental factors and levels

Factor	Low (−1)	High (+1)
Plunger shape	Rectangular	Circular
Plunger size	6.4 cm	9.7 cm
Type of surface	Concentrated	Distributed
Plunger force	0 N	40 N
Rotation	None	Rotation

A Factorial Problem

Otto and Wood (2001) describe the use of a factorial experiment to aid in the design of a household blender. The objective is to design a device that will liquefy solid vegetable matter. The experimenters chose to attempt to blend 200 g of vegetables for 10 s. The response variable is the percent of the 200 g that are liquid after 10 s of blending. Their design concept involves adding a plunger that imparts a force on the vegetables as they are being blended. There were five factors:

1. Plunger shape
2. Plunger size
3. Type of surface (part of the plunger that contacts the vegetables)
4. Plunger force (force imparted on the vegetables)
5. Plunger rotation

Suppose the experimenters decided to use a one-half fraction of a 2^5 design, that is, 2^{5-1}, which is resolution V. That is, they decided to use exactly two different values for each factor. Table 9.1 shows the factors and their levels.

Figure 9.4 shows the experimental design alias structure as generated using Minitab 16.

Notice that the generator effect for this design is the five-way cross-product term, so that each main effect is aliased with a four-way cross-product and each two-way cross-product is aliased with a

5/12/2014 7:35:34 AM

Welcome to Minitab, press F1 for help.

Fractional Factorial Design

Factors: 5 Base Design: 5, 16 Resolution: V

Runs: 16 Replicates: 1 Fraction: 1/2

Blocks: 1 Center pts (total): 0

Design Generators: E = ABCD

Alias Structure

I + ABCDE

A + BCDE

B + ACDE

C + ABDE

D + ABCE

E + ABCD

AB + CDE

AC + BDE

AD + BCE

AE + BCD

BC + ADE

BD + ACE

BE + ACD

CD + ABE

CE + ABD

DE + ABC

Fig. 9.4 Blender experimental design

Table 9.2 Runs for the 2^{5-1} blender experiment

Run	P.Shape	P.Size	Surface	P.Force	Rotation
1	−1	−1	−1	−1	1
2	1	−1	−1	−1	−1
3	−1	1	−1	−1	−1
4	1	1	−1	−1	1
5	−1	−1	1	−1	−1
6	1	−1	1	−1	1
7	−1	1	1	−1	1
8	1	1	1	−1	−1
9	−1	−1	−1	1	−1
10	1	−1	−1	1	1
11	−1	1	−1	1	1
12	1	1	−1	1	−1
13	−1	−1	1	1	1
14	1	−1	1	1	−1
15	−1	1	1	1	−1
16	1	1	1	1	1

three-way cross-product. The model we will fit to data will therefore only include main effects and two-way cross-product terms.

Table 9.2 shows the runs for this ½ fractional design.

Each run was replicated twice (each of the 16 prototypes were run with two loads of 200 g each). The data are given in Table 9.3.

Figure 9.5 shows the ANOVA output from JMP 11.

Figure 9.6 shows the residual plot for this model. Inasmuch as the residuals plotted against the expected results average percent blended for each run) show no discernable pattern, it seems reasonable to believe that the underlying assumption of constant noise variance is true. Thus, the p-values are fair indicators that the effects we have observed are repeatable. All main effects were significant ($p < 0.0001$). Table 9.4 shows the average response for each main effect at each of the two levels. Three two-way cross-products were significant; P.Shape*Rotation, P.Size*P.Force, and P. Force*Rotation. Figure 9.7a, b, and c shows the interaction plots for the significant cross-products.

If the experimenters were to use the mean results by factor and level, then they might conclude that the best design would be P.Shape = circular, P.Size = 9.7 cm (radius of plunger), Surface = concentrated, P.Force = 40 N, and Rotation = Rotate. There are some possible problems with this conclusion. The interaction plot of P.Force*Rotation indicates that P.Force = 0 N and Rotation = None is the best (i.e., highest percent blended). A post-hoc comparison indicates that in fact there is a significant (repeatable) difference between all four combinations of P.Force and Rotation, and that P. Force = 40 N and Rotation = None is the best configuration. Figure 9.8 shows the comparisons, together with p-values, using the Tukey Honestly Significant Difference (HSD) method.

Finally, consider the run that yielded the best results, namely run 12, which had P.Shape = circular, P.Size = 9.7 cm, Surface = concentrated, P.Force = 40 N, and Rotation = None. The data seem to lead to three different conclusions about the best design. Of the three contenders, only one has data (Run 12) that were actually observed in the experiment. One way to adjudicate this situation is to compute predicted values for each of the three possible configurations. A model can be fit, using the Helmert-coded levels, and incorporating only those interactions that were significant in the ANOVA. Figure 9.9 shows the regression fit output. Note that all coefficients are significantly different from 0 ($p < 0.0001$), and adjusted R^2 is approximately 0.9963, indicating a good fit. Table 9.5 summarizes the three possible configurations, together with predicted response values and associated 95 % confidence intervals.

Table 9.3 Blender experiment data

Run	P.Shape	P.Size (cm)	Surface	P.Force (N)	Rotation	Percent Blended
1	Rect.	6.4	Concentrated	0	Rotate	63.86
1	Rect.	6.4	Concentrated	0	Rotate	66.91
2	Circ.	6.4	Concentrated	0	None	42.28
2	Circ.	6.4	Concentrated	0	None	41.05
3	Rect.	9.7	Concentrated	0	None	7.80
3	Rect.	9.7	Concentrated	0	None	6.22
4	Circ.	9.7	Concentrated	0	Rotate	67.38
4	Circ.	9.7	Concentrated	0	Rotate	68.57
5	Rect.	6.4	Distributed	0	None	3.73
5	Rect.	6.4	Distributed	0	None	2.58
6	Circ.	6.4	Distributed	0	Rotate	64.86
6	Circ.	6.4	Distributed	0	Rotate	65.14
7	Rect.	9.7	Distributed	0	Rotate	44.82
7	Rect.	9.7	Distributed	0	Rotate	47.04
8	Circ.	9.7	Distributed	0	None	24.22
8	Circ.	9.7	Distributed	0	None	26.47
9	Rect.	6.4	Concentrated	40	None	29.32
9	Rect.	6.4	Concentrated	40	None	24.64
10	Circ.	6.4	Concentrated	40	Rotate	36.39
10	Circ.	6.4	Concentrated	40	Rotate	40.36
11	Rect.	9.7	Concentrated	40	Rotate	70.86
11	Rect.	9.7	Concentrated	40	Rotate	68.62
12	Circ.	9.7	Concentrated	40	None	94.66
12	Circ.	9.7	Concentrated	40	None	95.84
13	Rect.	6.4	Distributed	40	Rotate	20.16
13	Rect.	6.4	Distributed	40	Rotate	20.22
14	Circ.	6.4	Distributed	40	None	45.42
14	Circ.	6.4	Distributed	40	None	44.70
15	Rect.	9.7	Distributed	40	None	56.00
15	Rect.	9.7	Distributed	40	None	54.08
16	Circ.	9.7	Distributed	40	Rotate	69.80
16	Circ.	9.7	Distributed	40	Rotate	67.40

It appears that in this case, the run with the best observed response values (Run 12, Alt. 3) is the optimal choice for the product design. The presence of interaction effects made the choice indicated by maximizing the average results of the main effects (Alt. 1) yield a suboptimal design choice. Furthermore, simply choosing the design based on the optimal results for only one of the significant interaction effects also proved to yield a less desirable alternative (Alt. 2).

Despite the fact that the conditions of Run 12 were clearly the best, based on the model and experimental data, it would be prudent to obtain some additional data with all three alternatives. Table 9.6 shows the results of three replicates of the experiment with the three alternative designs.

The means and standard deviations (SD) of percent blended for each of the alternatives is given in Table 9.7.

The confirmatory experiment indicates that in fact Alternative 3 (Run 12 from original experiment) appears to be the winning design for the product. It also appears to have the added benefit of having the most consistent results (smallest SD). In order to test whether we should believe that the reduced SD is a repeatable phenomenon, we can compute the ratios of variances:

Response Percent Blended

Summary of Fit

RSquare	0.997937
RSquare Adj	0.996002
Root Mean Square Error	1.582415
Mean of Response	46.29311
Observations (or Sum Wgts)	32

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	15	19377.324	1291.82	515.8955
Error	16	40.065	2.50	Prob > F
C. Total	31	19417.388		<.0001*

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
P.Shape	1	1	2958.9206	1181.660	<.0001*
P.Size	1	1	2082.5355	831.6712	<.0001*
Surface	1	1	883.2834	352.7437	<.0001*
P.Force	1	1	1194.7091	477.1131	<.0001*
Rotation	1	1	2509.4206	1002.150	<.0001*
P.Shape*P.Size	1	1	3.2101	1.2820	0.2742
P.Shape*Surface	1	1	3.8355	1.5317	0.2337
P.Shape*P.Force	1	1	1.2508	0.4995	0.4899
P.Shape*Rotation	1	1	730.2709	291.6374	<.0001*
P.Size*Surface	1	1	4.5860	1.8314	0.1948
P.Size*P.Force	1	1	4370.2315	1745.274	<.0001*
P.Size*Rotation	1	1	0.7776	0.3105	0.5851
Surface*P.Force	1	1	0.1652	0.0660	0.8006
Surface*Rotation	1	1	0.0364	0.0145	0.9055
P.Force*Rotation	1	1	4634.0909	1850.648	<.0001*

Fig. 9.5 ANOVA output—blender experiment

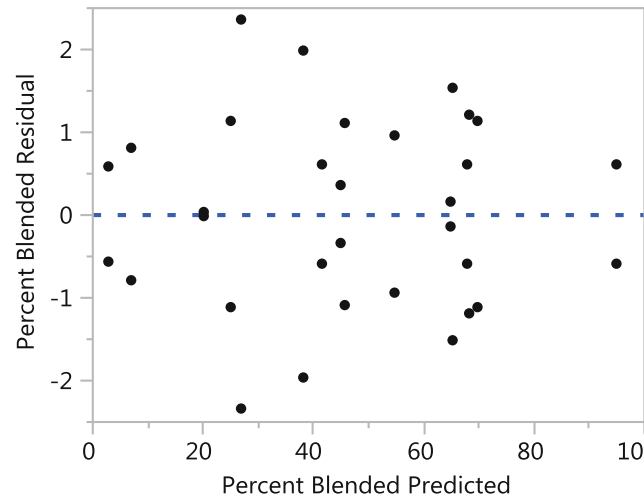


Fig. 9.6 Residual by expected plot

Table 9.4 Mean response by factor and level

Factor	Level	Least sq mean	Std error
P.Shape	Rect.	36.677	0.3956
P.Shape	Circ.	55.909	0.3956
P.Size	6.4 cm	38.226	0.3956
P.Size	9.7 cm	54.360	0.3956
Surface	Concentrated	51.547	0.3956
Surface	Distributed	41.039	0.3956
P.Force	0 N	40.183	0.3956
P.Force	40 N	52.403	0.3956
Rotation	None	37.438	0.3956
Rotation	Rotate	55.149	0.3956

$$F_1 = \frac{SD^2_{Alt.1}}{SD^2_{Alt.2}} \approx \frac{1.363^2}{0.319^2} \approx 18.26$$

Under the null hypothesis of no difference in variances between Alternative 1 and Alternative 2, the statistic F_1 has an F distribution with $3 - 1 = 2^\circ$ of freedom in both numerator and denominator. The p-value from an F distribution with 2° of freedom in the numerator and denominator is approximately 0.0519, which is close to significant. Similarly,

$$F_2 = \frac{SD^2_{Alt.2}}{SD^2_{Alt.3}} \approx \frac{0.586^2}{0.319^2} \approx 3.37$$

which has a p-value of approximately 0.2288. This is clearly not significant. Although a larger sample size might yield significant p-values, the experimenters/designers should determine whether the additional experimentation would be worth knowing that the most desirable alternative in mean response is also the most consistent.

Another Factorial Problem

From elementary physics, we know that voltage, V, is electrical potential, the potential energy required to move electrical charge, q, through a conductor. The change in voltage is electromotive force, namely the force required to move a charge through the conductor between two points along the conductive path. Work, W, is the change in voltage times the electrical charge. Power, P, is the work per unit time, t. The mathematical expressions are the following:

$$\Delta V = \frac{W}{q} => W = q\Delta V$$

$$P = \frac{W}{t} = \frac{q\Delta V}{t}$$

From Ohm's law, we know the relationship between voltage, V, current, I, and resistance, R:

$$V = IR$$

So:

$$\Delta V = IR - I_0 R_0$$

where I_0 and R_0 are the initial values of current and resistance.

An implantable defibrillator is a device that is used to "restart" a heart that has stopped beating. The most commonly known cause of heart stoppage is called ventricular fibrillation. Then ventricles

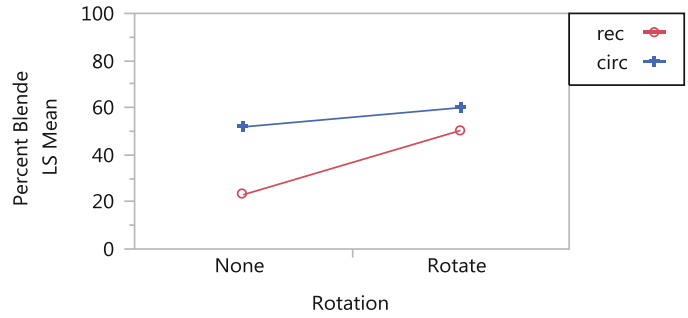
a

P.Shape*Rotation

Least Squares Means Table

Level	Least Sq Mean	Std Error
rect.,None	23.044572	0.55946819
rect.,Rotate	50.309776	0.55946819
circ.,None	51.830708	0.55946819
circ.,Rotate	59.987394	0.55946819

LS Means Plot



b

P.Size*P.Force

Least Squares Means Table

Level	Least Sq Mean	Std Error
6.4cm,0N	43.802038	0.55946819
6.4cm,40N	32.649846	0.55946819
9.7cm,0N	36.563768	0.55946819
9.7cm,40N	72.156797	0.55946819

LS Means Plot

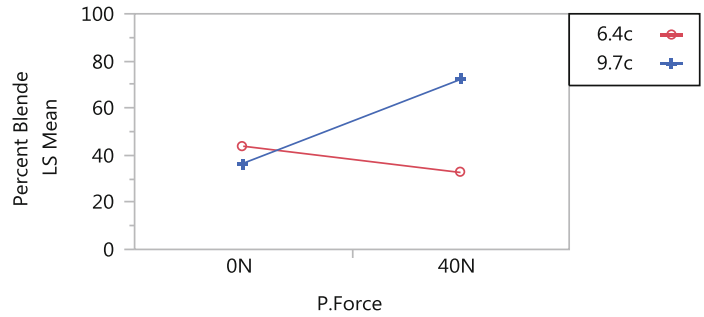


Fig. 9.7 (**a**) Significant interaction plots—P.Shape*rotation; (**b**) significant interaction plots—P.Size*P.Force; (**c**) significant interaction plots—P.Force*rotation

C

P.Force*Rotation

Least Squares Means Table

Level	Least Sq Mean	Std Error
0N,None	19.293506	0.55946819
0N,Rotate	61.072300	0.55946819
40N,None	55.581774	0.55946819
40N,Rotate	49.224870	0.55946819

LS Means Plot

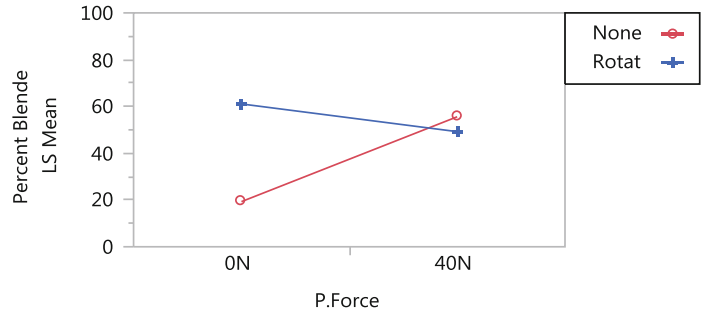


Fig. 9.7 (continued)

P.Force*Rotation

Least Squares Means Table

Level	Least Sq Mean	Std Error
0N,None	19.293506	0.55946819
0N,Rotate	61.072300	0.55946819
40N,None	55.581774	0.55946819
40N,Rotate	49.224870	0.55946819

LSMeans Differences Tukey HSD

α=
0.050 Q=
2.86102

Level				Least Sq Mean
0N,Rotate	A			61.072300
40N,None		B		55.581774
40N,Rotate			C	49.224870
0N,None			D	19.293506

Levels not connected by same letter are significantly different.

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
0N,Rotate	0N,None	41.77879	0.7912075	39.51513	44.04245	<.0001*
40N,None	0N,None	36.28827	0.7912075	34.02461	38.55193	<.0001*
40N,Rotate	0N,None	29.93136	0.7912075	27.66770	32.19502	<.0001*
0N,Rotate	40N,Rotate	11.84743	0.7912075	9.58377	14.11109	<.0001*
40N,None	40N,Rotate	6.35690	0.7912075	4.09324	8.62056	<.0001*
0N,Rotate	40N,None	5.49053	0.7912075	3.22687	7.75419	<.0001*

Fig. 9.8 Pairwise comparisons of P.Force*rotation cross-product (tukey HSD)

Response Percent Blended

Summary of Fit

RSquare	0.997223
RSquare Adj	0.996257
Root Mean Square Error	1.531213
Mean of Response	46.29311
Observations (or Sum Wgts)	32

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	8	19363.462	2420.43	1032.338
Error	23	53.926	2.34	**Prob > F**
C. Total	31	19417.388		<.0001*

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	7	13.861504	1.98021	0.7908
Pure Error	16	40.064596	2.50404	**Prob > F**
Total Error	23	53.926100		0.6055
				Max RSq

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	46.293113	0.270683	171.02	<.0001*
P.Shape	9.6159383	0.270683	35.52	<.0001*
P.Size	8.0671702	0.270683	29.80	<.0001*
Surface	-5.253818	0.270683	-19.41	<.0001*
P.Force	6.1102094	0.270683	22.57	<.0001*
Rotation	8.8554725	0.270683	32.72	<.0001*
P.Shape*Rotation	-4.777129	0.270683	-17.65	<.0001*
P.Size*P.Force	11.686305	0.270683	43.17	<.0001*
P.Force*Rotation	-12.03392	0.270683	-44.46	<.0001*

Fig. 9.9 Multiple regression Fit of reduced model—helmert-coded regressors

Table 9.5 Possible conclusions about optimal design

Factor	Alt. 1 Main effects	Alt. 2 Interaction	Alt. 3 Best run
P.Shape	Circular (+1)	Circular (+1)	Circular (+1)
P.Size	9.7 cm (+1)	9.7 cm (+1)	9.7 cm (+1)
Surface	Concentrated (−1)	Concentrated (−1)	Concentrated (−1)
P.Force	40 N (+1)	0 N (−1)	40 N (+1)
Rotation	Rotate (+1)	None (−1)	None (−1)
Predicted response	79.07	35.32	**94.98**
LCL (mean)	77.39	33.64	**93.30**
UCL (mean)	80.75	37.00	**96.66**

Table 9.6 Three alternative designs: Confirmatory experiment results

Alternative	P.Shape	P.Size (cm)	Surface	P.Force (N)	Rotation	Percent blended
Alt. 1	Circ.	9.7	Concentrated	40	Rotate	80.84
Alt. 1	Circ.	9.7	Concentrated	40	Rotate	78.65
Alt. 1	Circ.	9.7	Concentrated	40	Rotate	81.16
Alt. 2	Circ.	9.7	Concentrated	0	None	35.69
Alt. 2	Circ.	9.7	Concentrated	0	None	36.82
Alt. 2	Circ.	9.7	Concentrated	0	None	35.99
Alt. 3	Circ.	9.7	Concentrated	40	None	96.15
Alt. 3	Circ.	9.7	Concentrated	40	None	96.34
Alt. 3	Circ.	9.7	Concentrated	40	None	95.71

Table 9.7 Means and standard deviations of percent blended for alternative designs

Alternative	N	Mean % blended	SD
Alt. 1	3	80.216	1.363
Alt. 2	3	36.165	0.586
Alt. 3	3	96.067	0.319

are the power pumps of the heart. They regularly pump blood into and out of the heart, and their pumping cycle is regulated by electrical impulses. If those impulses are interrupted or altered, the ventricles may not pump. If the pulses increase their frequency, the ventricles may pump so rapidly that they fail. Rather than pumping steadily, they may simply vibrate, or fibrillate. Ventricular fibrillation (VF) is a form of heart attack.

An implantable defibrillator has a sensor that detects increased frequency of the electrical impulses being delivered to the ventricles. If the rate increases beyond some predetermined threshold, the defibrillator will discharge electrical energy to the heart's regulator (usually in the range of 200–360 J) to hopefully interrupt the increased impulse rate, and avoid ventricular fibrillation.

There are several factors that affect the delivered voltage. There is the resistance of the cardiac lead (wire cable) that delivers the voltage. There is the voltage source (battery). The resistance encountered at the site where the lead contacts the heart's natural pacemaker affects the delivered voltage. The power required to end fibrillation is a probabilistic quantity. That is, a given power level has some probability of stopping the VF. Even if total resistance is known, and the voltage from the battery is fixed, the required power cannot be known with certainty.

The energy unit Joules (J) can be expressed as coulomb-volts:

$$J = CV$$

Coulombs, in turn, are a product of current, in amperes, and duration of electrical flow, in seconds, s. Thus:

$$C = Is$$
$$J = IVs$$

From Ohm's law, current is the ratio of voltage to resistance:

$$I = \frac{V}{R}$$

So, the energy (E, measured in Joules) delivered by the defibrillator is:

$$E = \frac{V^2}{R} s$$

In addition to controlling V and s, the EAS may choose to use either a monophasic or biphasic wave form. That means either current flows in one direction, or current changes direction in the middle of the wave form. Generally, it has been shown that biphasic waveforms require less energy E to cardiovert (defibrillate) than the monophasic approach. The voltage and the duration of the shock can be controlled. The resistance, R, is a function of the patient's physiology, so it cannot be controlled. However, it can be measured.

In practice, the energy to charge the capacitor in the defibrillator is selected. The voltage can be calculated:

$$V = \sqrt{\frac{ER}{s}}$$

Suppose that the EAS has a cardiac simulator, where heart tissue resistance (in ohms) can be selected, and where VF can be simulated. This simulator allows the EAS to simulate the action of the defibrillator, which may or may not stop the VF event. The EAS decides that the factors to be varied, and the ranges over which they will be varied are:

Energy (E)—in Joules—180–360 J (center value = 270 J)
Heart Resistance (R)—in ohms—2250–5400 Ω (center = 3825 Ω)
Wave-form duration (s)—in seconds—0.02–0.03 s (center = 0.025 s)

The EAS decides to use a 3-factor CCD experiment. In each "run", the experiment will be replicated 100 times. The response for the experiment will be the proportion of times in which defibrillation was achieved. Since the simulator includes some random components, so each time the simulator is run, the outcome is uncertain. The experiment will be repeated at the centerpoint run two times, to get an estimate of variability.

Table 9.8 shows the design points in natural and coded units, and the response (Percent Success).

Figure 9.10 shows the JMP output for fitting the full second-order model.

Five of the terms have coefficients that are not significantly different from zero. One of them is a main effect, s, the duration of the pulse. Three of them are higher-order terms with s as a component. Apparently the pulse duration, as long as it is between 20 and 30 ms, has no appreciable effect on the probability of defibrillating. Refitting the model without these terms will yield a more parsimonious (remember our friend, William Ockham) model. The fit is shown in Figure 9.11.

Now all terms are significant. The final model, with regressors in Helmert-coded units, is:

$$\Pr\{\text{defibrillation}\} = 93.03 + 12.42*E - 7.57*R + 6.84*E*R - 7.32*E^2$$

Recall that Pr{defibrillation} is expressed as a percent.

Since resistance, R, is not a design variable, one can ask the question, "for a given value of R, what value of E will make Pr{defibrillation} $\geq$ 95 %, subject to the constraints that $E \leq$ 360 J ?" Table 9.9 shows some results.

Of course, 95 % may not be a sufficient likelihood for achieving defibrillation. However, the point is that the model can be used to make a guess at the energy required. The EAS can then use the fundamental relationships:

Table 9.8 Central composite design with percent success data

E	R	s	E-coded	R-coded	s-coded	Percent success (%)
180	2250	0.02	−1	−1	−1	79
180	2250	0.03	−1	−1	1	78
180	3825	0.025	−1	0	0	89
180	5400	0.02	−1	1	−1	95
180	5400	0.03	−1	1	1	94
270	2250	0.025	0	−1	0	80
270	3825	0.02	0	0	−1	90
270	3825	0.025	0	0	0	90
270	3825	0.025	0	0	0	90
270	3825	0.03	0	0	1	89
270	5400	0.025	0	1	0	95
360	2250	0.02	1	−1	−1	82
360	2250	0.03	1	−1	1	81
360	3825	0.025	1	0	0	91
360	5400	0.02	1	1	−1	96
360	5400	0.03	1	1	1	95

Response Percent Success

Summary of Fit

RSquare	0.99442
RSquare Adj	0.98605
Root Mean Square Error	1.589104
Mean of Response	88.45292
Observations (or Sum Wgts)	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	9	2700.1937	300.022	118.8086
Error	6	15.1515	2.525	**Prob > F**
C. Total	15	2715.3452		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	93.565317	0.752333	124.37	<.0001*
E-coded	12.418758	0.502519	24.71	<.0001*
R-coded	-7.56774	0.502519	-15.06	<.0001*
s-coded	-0.295609	0.502519	-0.59	0.5778
E-coded*R-coded	6.8446563	0.561833	12.18	<.0001*
E-coded*s-coded	0.1192137	0.561833	0.21	0.8390
R-coded*s-coded	-0.357064	0.561833	-0.64	0.5485
E-coded*E-coded	-6.569138	0.9787	-6.71	0.0005*
R-coded*R-coded	-1.585718	0.9787	-1.62	0.1563
s-coded*s-coded	-0.024973	0.9787	-0.03	0.9805

Fig. 9.10 Defibrillator full second-order model

Response Percent Success

Summary of Fit

RSquare	0.990936
RSquare Adj	0.987641
Root Mean Square Error	1.495771
Mean of Response	88.45292
Observations (or Sum Wgts)	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	2690.7345	672.684	300.6636
Error	11	24.6106	2.237	Prob > F
C. Total	15	2715.3452		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	93.02842	0.610646	152.34	<.0001*
E-coded	12.418758	0.473004	26.26	<.0001*
R-coded	-7.56774	0.473004	-16.00	<.0001*
E-coded*R-coded	6.8446563	0.528835	12.94	<.0001*
E-coded*E-coded	-7.320794	0.772413	-9.48	<.0001*

Fig. 9.11 Defibrillator reduced second-order model

Table 9.9 Model results for achieving Pr{defibrillation} $\geq$ 95 %

E (coded)	E (J)	R (coded)	R (Ω)	Pr{Defib} (%)
−0.572	218.5	−1	2250	95.01
0.177	285.9	0	3825	95.00
0.662	329.6	1	5400	95.00

$$E = IVs$$
$$I = \frac{V}{R}$$
$$V = \sqrt{\frac{ER}{s}}$$

to determine voltage and current.

Key Points

- Applications of experimental designs include range finding and optimization.
- Confidence intervals for predicted values will provide a means incorporating variability in product/system design.

Exercises and Questions

Design an experiment, execute it, and fit a model/analyze the data. Discuss why you chose the particular design and the model.

Chapter 10
Binary Logistic Regression

What Are the Odds?

Most people have heard the term "Odds", and know that it has something to do with the likelihood of obtaining a "successful" outcome in some sort of game or trial. If P represents the probability of "success", then $1 - P$ is the probability of "not success". The odds are:

$$O = \frac{P}{1-P}$$

As an example, suppose the "game" is to toss a 6-sided die. Perhaps we are interested in the odds of obtaining an even number of dots face up after the throw. Clearly there are $n = 6$ possibilities for the outcome, and 3 of them (2, 4, 6) are even. So:

$$P = \frac{3}{6}$$

And

$$1 - P = \frac{3}{6}$$

The odds of an even result (in comparison to not even) is then:

$$O = \frac{P}{1-P} = \frac{3/6}{3/6} = \frac{3}{3} = 1$$

Thus, the odds of "even" versus "not even" are 1, which, ironically, is called "even odds". That is, there is no greater or lesser likelihood of obtaining an "even" number versus a "not even" number. Suppose we consider a different pair of events, namely obtaining either a "1", "2", "3" or a "5" versus anything else (either "4" or "6"). Then we have:

$$P = \frac{4}{6}$$

S.A. Pardo, *Empirical Modeling and Data Analysis for Engineers and Applied Scientists*,
DOI 10.1007/978-3-319-32768-6_10

$$1 - P = \frac{2}{6}$$

$$O = \frac{P}{1-P} = \frac{4/6}{2/6} = \frac{4}{2} = 2$$

Now we say that it is two times as likely for obtaining either "1", "2", "3" or a "5" versus not, or the odds of either "1", "2", "3" or a "5" are 2 to 1 in favor. Now, if you are told that the odds are 3 to 1 in favor of a particular horse to win a race, you know that it is three times as likely that this horse will win, versus not (i.e., some other horse will win).

An odds ratio is the ratio of odds for two different conditions. Suppose there are two teams in a sporting event. Each team has some probability of winning, and consequently a probability of losing. Let P_1 represent the probability that team 1 will win, and P_2 the win probability for team 2. We presume that the probabilities of winning are not conditioned on the fact that these particular two teams are competing against each other. Thus P_1 and P_2 are not necessarily complementary; P_1 is not necessarily equal to $1 - P_2$. Then the odds for each team are:

$$O_1 = \frac{P_1}{1-P_1}$$

$$O_2 = \frac{P_2}{1-P_2}$$

The odds ratio is then:

$$\frac{O_1}{O_2} = \frac{P_1/(1-P_1)}{P_2/(1-P_2)}$$

The odds ratio in the number of times more likely the "numerator" team is to win compared to the "denominator" team. Suppose $P_1 = 0.90$ and $P_2 = 0.45$. Then the odds ratio of team 1 to team 2 is:

$$\frac{O_1}{O_2} = \frac{P_1/(1-P_1)}{P_2/(1-P_2)} = \frac{0.90/0.10}{0.45/0.65} \approx 13$$

So team 1 is 13 times more likely than team 2 to win the contest. Note that the ratio of P_1 to P_2 is only $0.90/0.45 = 2$. This ratio does not incorporate the probability that a given team could lose.

The Logit Transformation

It will be convenient to use the natural logarithm of odds rather than the odds directly. So we will define the logit (pronounced "low-jit") of P to be:

$$\lambda = \ln(O) = ln\left(\frac{P}{1-P}\right)$$

If one was given the logit's value, one could solve for P:

$$P = \frac{1}{1 + e^{-\lambda}}$$

Keep in mind that P represents the probability of obtaining a "success", however success is defined. Furthermore, we will define a Bernoulli random variable, Y, to have the following binary values:

$$Y = \begin{cases} 1 & \text{if "success"} \\ 0 & \text{otherwise} \end{cases}$$

Thus,

$$P = \Pr\{Y = 1\}$$

Suppose that the logit was in fact a linear function of some regressor, X, so that:

$$\lambda(X) = \ln(O) = ln\left(\frac{P(X)}{1 - P(X)}\right) = \beta_0 + \beta_1 X$$

By solving for $P(X)$, we get:

$$P(X) = \Pr\{Y = 1|X\} = \frac{1}{1 + e^{-\lambda(X)}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

Once data are collected, the parameters β_0 and β_1 may be estimated. The estimates may be obtained in many ways, but a common way is via maximum likelihood. The likelihood function for a sample of binary results, $y_1, y_2, \ldots, y_n$, and the associated regressor values $x_1, x_2, \ldots, x_n$, is given by:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \pi^{y_i}(x_i)(1 - \pi(x_i))^{1-y_i}$$

where:

$$\pi(x_i) = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} = \frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} = Pr\{Y = 1|x_i\}$$

The idea is to find values of β_0 and β_1 that maximize the likelihood function. Generally it is easier to maximize the logarithm of the likelihood function, since it involves sums and not products. Taking the natural log of $L(\boldsymbol{\beta})$, differentiating with respect to β_0 and β_1, and setting the partial derivatives equal to 0 gives the equations:

$$\sum_{i=1}^{n} (y_i - \pi(x_i)) = 0$$

and

$$\sum_{i=1}^{n} x_i(y_i - \pi(x_i)) = 0$$

Inasmuch as both of these equations are non-linear with respect to the two unknown parameters, there is no closed form solution. Rather, numerical approximation methods such as Newton-Raphson must be employed.

Once the model parameters are estimated, a measure of goodness is given by:

$$D = -2\sum_{i=1}^{n}\left[y_i ln\left(\frac{\widehat{\pi}_i}{y_i}\right) + (1 - y_i) ln\left(\frac{1 - \widehat{\pi}_i}{1 - y_i}\right)\right]$$

Where $\widehat{\pi}_i = \frac{1}{1+e^{\hat{\beta}_0+\hat{\beta}_1 x_i}}$ is the predicted probability that $Y = 1$ when $X = x_i$. Note that if $y_i = 0$, then the first term in the sum is set to 0, and if $y_i = 1$, then the second term is set to 0. The quantity D is called the deviance (Hosmer and Lemeshow 1989), and is analogous to the sums of squares for error in a usual multiple regression model. It can be used to compare the goodness of fit for models with different combinations of regressors. In general, lower deviance is more desirable.

It is fairly easy to generalize the logistic equation to multiple regressors:

$$P(X) = \Pr\{Y = 1|\boldsymbol{x}\} = \frac{1}{1 + e^{-\lambda(\boldsymbol{x})}} = \frac{1}{1 + e^{-\boldsymbol{x}'\boldsymbol{\beta}}}$$

$\boldsymbol{x}' = [1\ x_1\ x_2 \cdots x_k]$ is a vector of regressors (the "1" is the "regressor" for the intercept) and

$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}$ is a vector of unknown parameters.

The values of the x_j could be chosen in a designed experimental fashion, say a 2^{k-p} design. The y_i are binary observations. For the Moment, we will assume that the x_j are continuously valued regressors, and our objective is to find a predictive equation for the probability of a "success" ($Y = 1$) in terms of the x_j. Furthermore, in building this probability model, we would like to determine which if any of the regressors (x_j) actually affect the probability that $Y = 1$.

Example: Continuous Regressors

Consider the problem of designing a natural gas pipeline (White 2012). Suppose that the EAS wants to maximize the probability that the pipe will not leak. Furthermore, to simplify the problem, suppose that there are only three design factors to be considered:

Pipe outer diameter (OD)
Pipe wall thickness (WT)
Specified minimum yield strength (SMYS)

OD and WT are fairly obvious in their nature, and will be measured in inches. SMYS is defined to be the force (in psi) at which the steel used to make the pipe begins to stretch (White, ref.cit.). Table 10.1 shows these factors and the range over which our EAS is interested.

The EAS asks the engineering lab to put together several prototype systems, and perform a leak test. The response for each test is binary; either the pipe leaked ($Y = 0$) or not ($Y = 1$). Furthermore,

Table 10.1 Pipeline factors and levels

Factor	Low	Middle	High	Units
Outer diameter (OD)	48	60	72	Inches
Wall thickness (WT)	6	12	18	Inches
Specified max yield strength (SMYS)	60,000	70,000	80,000	psi

Table 10.2 Box-Behnken design—3 factors, 13 runs

Run	OD	WT	SMYS
1	48	6	70,000
2	48	12	60,000
3	48	12	80,000
4	48	18	70,000
5	60	6	60,000
6	60	6	80,000
7	**60**	**12**	**70,000**
8	60	18	60,000
9	60	18	80,000
10	72	6	70,000
11	72	12	60,000
12	72	12	80,000
13	72	18	70,000

the EAS decides to fit a logistic function using data from a 3-factor Box-Behnken design, with each design point replicated n = 15 times. Table 10.2 shows the design points (runs). Run #7 is the center point.

The EAS first uses Helmert coding to transform the regressors into the interval (−1,+1), and then fits the data to the model using the R code shown in Fig. 10.1. The R output is shown in Fig. 10.2.

The last line of code stores all the predicted probabilities for "success" (no leak), as computed by the logistic model. Figure 10.3 shows SAS code for fitting the same model. The SAS procedure GENMOD allows the use to compute confidence intervals for predicted values at each point. Figure 10.4 shows the SAS output. Table 10.3 shows the predicted probabilities of No Leak, together with 95 % confidence limits (LCL, UCL) for each run.

One might ask just how the confidence intervals for predicted values are computed. Assume the vector:

$$\boldsymbol{x} = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_k \end{bmatrix}$$

represents a particular point in the regressor space. Then the predicted logit at this point is:

$$\widehat{\lambda}(\boldsymbol{x}) = \boldsymbol{x}'\widehat{\boldsymbol{\beta}}$$

and

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\beta}_0 \\ \vdots \\ \widehat{\beta}_k \end{bmatrix}$$

is the vector of coefficient estimates.

```
setwd("H:\\Personal Data\\Experimentation for Design & Validation\\Data & Analyses\\")
df1 <- read.csv("20140813_pipeline.csv")
attach(df1)
model_bin <- glm(No.Leak ~ OD.coded + WT.coded + SMYS.coded,
+ family = binomial("logit"),na.action=na.omit)
summary(model_bin)
confint(model_bin)
anova(model_bin,test="Chisq") #computes a sequential likelihood ratio chi squared #statistic for
each effect in the model
#
df1$probmod <- predict(model_bin,type="response")
#
# for glm objects the predict function defaults to returning log odds ratios
# the parameter type ="response" will make predict return predicted probabilities of a "1"
response
#
tapply(probmod,SMYS.coded,quantile)
tapply(probmod,WT.coded,quantile)
tapply(probmod,OD.coded,quantile)
write.csv(df1,file="20140814.pipeline.probs.csv")
```

Fig. 10.1 Logistic regression R code

The variance of the estimated logit at $\boldsymbol{x}$ is:

$$V\left(\widehat{\lambda}\right) = \boldsymbol{x}'\widehat{\boldsymbol{\Sigma}}\boldsymbol{x}$$

where $\widehat{\boldsymbol{\Sigma}}$ is the estimate of the variance-covariance matrix of the coefficient estimates. The standard error, $SE\left(\widehat{\lambda}\right)$ of the estimated logit is the square root of the variance.

The confidence limits for the logit are:

$$\widehat{\lambda}(\boldsymbol{x}) \pm SE\left(\widehat{\lambda}(\boldsymbol{x})\right) = \boldsymbol{x}'\widehat{\boldsymbol{\beta}} \pm z_{1-\frac{\alpha}{2}}\sqrt{\boldsymbol{x}'\widehat{\boldsymbol{\Sigma}}\boldsymbol{x}}$$

The computation of the estimated variance-covariance matrix is given by Hosmer and Lemeshow (1989). The variance-covariance matrix is the inverse of the information matrix. The information matrix has diagonal elements:

$$\sum_{i=1}^{n} x_{ij}^2 p_i(1-p_i),$$

where the sum is over all the individual observations, and the off-diagonal elements are:

$$\sum_{i=1}^{n} x_{ij}x_{il}p_i(1-p_i)$$

```
setwd("H:\\Personal Data\\Experimentation for Design & Validation\\Data &
Analyses\\")
> df1 <- read.csv("20140813_pipeline.csv")
>
> attach(df1)
> model_bin <- glm(No.Leak ~ OD.coded + WT.coded + SMYS.coded,family =
binomial("logit"),na.action=na.omit)
>
> summary(model_bin)

Call:
glm(formula = No.Leak ~ OD.coded + WT.coded + SMYS.coded, family =
binomial("logit"),
    na.action = na.omit)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.77081  -0.50964   0.07307   0.44988   2.05206

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.1320     0.2304   0.573  0.56659
OD.coded      1.5417     0.5345   2.884  0.00392 **
WT.coded      2.1432     0.5358   4.000  6.33e-05 ***
SMYS.coded    3.6493     0.5977   6.105  1.03e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 270.20  on 194  degrees of freedom
Residual deviance: 118.91  on 191  degrees of freedom
AIC: 126.91

Number of Fisher Scoring iterations: 7

> confint(model_bin)
Waiting for profiling to be done...
                  2.5 %      97.5 %
(Intercept)   -0.3189782   0.5910472
OD.coded       0.6907986   3.0069397
WT.coded       1.2935005   3.6102790
SMYS.coded     2.6647122   5.1900525
```

Fig. 10.2 Logistic regression R output

```
> anova(model_bin,test="Chisq") #computes a sequential likelihood ratio chi squared
statistic
Analysis of Deviance Table

Model: binomial, link: logit

Response: No.Leak

Terms added sequentially (first to last)


             Df Deviance Resid. Df  Resid.Dev  Pr(>Chi)
NULL                         194     270.20
OD.coded      1    5.682       193     264.52    0.01714 *
WT.coded      1   20.266       192     244.25    6.74e-06 ***
SMYS.coded    1  125.345       191     118.91    < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #                       for each effect in the model
> #
> df1$probmod <- predict(model_bin,type="response")
> #
> # for glm objects the predict function defaults to returning log odds ratios
> # the parameter type ="response" will make predict return predicted probabilities of a
"1" response
> #
>
> tapply(probmod,SMYS.coded,quantile)
$`-1`
       0%        25%        50%        75%       100%
0.003468696 0.005600981 0.064050325 0.141830938 0.201957029

$`0`
      0%       25%       50%       75%      100%
0.02784503 0.38474718 0.53296297 0.67558022 0.97847852

$`1`
     0%      25%      50%      75%     100%
0.8372891 0.8871400 0.9494515 0.9956931 0.9973342
```

Fig. 10.2 (continued)

```
> tapply(probmod,WT.coded,quantile)
$`-1`
      0%       25%       50%       75%      100%
0.003468696 0.021750946 0.206296105 0.497882656 0.837289081

$`0`
      0%       25%       50%       75%      100%
0.006311743 0.121788908 0.532962966 0.903757017 0.995146058

$`1`
     0%      25%      50%      75%     100%
0.2019570 0.5571744 0.8270294 0.9831924 0.9973342

> tapply(probmod,OD.coded,quantile)
$`-1`
      0%       25%       50%       75%      100%
0.006311743 0.022461707 0.351712622 0.732624416 0.903757017

$`0`
      0%       25%       50%       75%      100%
0.003468696 0.201957029 0.532962966 0.837289081 0.997334218

$`1`
     0%      25%      50%      75%     100%
0.1217889 0.3190076 0.6816129 0.9826454 0.9951461

>
> write.csv(df1,file="20140814.pipeline.probs.csv")
```

Fig. 10.2 (continued)

The p_i are the predicted probabilities at each observation:

$$p_i = \frac{1}{1 + e^{-\widehat{x_i'\beta}}}$$

The confidence intervals for the predicted values are given by inverting the logit transformation:

$$\frac{1}{1 + e^{-\left(\widehat{\lambda}(x) \pm SE\left(\widehat{\lambda}(x)\right)\right)}}$$

Fleiss et al. (2003) gives a simplified formula for the standard error of logits when there is a single regressor, X:

$$SE\left(logit\left(\widehat{P}(x)\right)\right) = \sqrt{se\left(\widehat{\beta}_0\right) + 2xcov\left(\widehat{\beta}_0, \widehat{\beta}_1\right) + x^2\left(se\left(\widehat{\beta}_1\right)\right)^2}$$

```
libname stuff 'H:\Personal Data\Experimentation for Design &
Validation\Data & Analyses';

data calc;

  set stuff.d20140815_pipeline; /* do not include the sas7bdat
extension here */
/*************************************************************
*                                                            *
* variables: x1 = OD_coded                                   *
*            x2 = WT_coded                                   *
*            x3 = SMYS_coded                                 *
*            y  = No_Leak                                    *
*                                                            *
*                                                            *
*                                                            *
*************************************************************/

  run;

proc sort data=calc;
  by OD_coded WT_coded SMYS_coded;
  run;

proc means data=calc;
  var No_Leak;
  run;



proc genmod data=calc descending; * the "descending" parameter makes
the model for Pr{Y = 1} ;
  model No_Leak = OD_coded WT_coded SMYS_coded/dist = bin link = logit;
  output out=NLoutest PREDICTED=predprob LOWER=lcl UPPER=ucl;
  run;

data stuff.o20140815_pipeline_output;

  set NLoutest;

  run;
```

Fig. 10.3 SAS proc GENMOD for logistic regression

where:

$$logit\left(\widehat{P}(x)\right) = \widehat{\beta}_0 + \widehat{\beta}_1 x$$

$$w_i = p_i(1 - p_i)$$

$$\overline{X}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

```
                         The SAS System         18:05 Friday, August 15, 2014   1

                       The MEANS Procedure

                  Analysis Variable : No_Leak

   N            Mean         Std Dev         Minimum         Maximum
ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
 195       0.5128205       0.5011222               0       1.0000000
ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ

                         The SAS System         18:05 Friday, August 15, 2014   2

                       The GENMOD Procedure

                         Model Information

                   Data Set                 WORK.CALC
                   Distribution              Binomial
                   Link Function                Logit
                   Dependent Variable         No_Leak


               Number of Observations Read          195
               Number of Observations Used          195
               Number of Events                     100
               Number of Trials                     195


                         Response Profile

                  Ordered                        Total
                    Value     No_Leak        Frequency

                        1     1                    100
                        2     0                     95

PROC GENMOD is modeling the probability that No_Leak='1'.


                Criteria For Assessing Goodness Of Fit

      Criterion                     DF           Value        Value/DF

      Log Likelihood                          -59.4532
      Full Log Likelihood                     -59.4532
      AIC (smaller is better)                 126.9063
      AICC (smaller is better)                127.1169
      BIC (smaller is better)                 139.9983


           Algorithm converged.


             Analysis Of Maximum Likelihood Parameter Estimates

                                Standard     Wald 95% Confidence          Wald
   Parameter     DF    Estimate     Error           Limits            Chi-Square    Pr > ChiSq

   Intercept      1      0.1320    0.2304     -0.3196      0.5836           0.33        0.5666
   OD_coded       1      1.5417    0.5345      0.4941      2.5893           8.32        0.0039
   WT_coded       1      2.1432    0.5358      1.0931      3.1933          16.00        <.0001
   SMYS_coded     1      3.6493    0.5978      2.4778      4.8209          37.27        <.0001
   Scale          0      1.0000    0.0000      1.0000      1.0000

NOTE: The scale parameter was held fixed.
```

Fig. 10.4 SAS proc GENMOD output

Table 10.3 Percentiles for predicted probability of no leak by run

Run	OD	WT	SMYS	Predicted prob	LCL	UCL
1	48	6	70,000	0.0278	0.0037	0.1790
2	48	12	60,000	0.0063	0.0008	0.0492
3	48	12	80,000	0.9038	0.7823	0.9608
4	48	18	70,000	0.6756	0.4800	0.8245
5	60	6	60,000	0.0035	0.0004	0.0282
6	60	6	80,000	0.8373	0.6763	0.9269
7	60	12	70,000	0.5330	0.4208	0.6419
8	60	18	60,000	0.2020	0.0961	0.3758
9	**60**	**18**	**80,000**	**0.9973**	**0.9774**	**0.9997**
10	72	6	70,000	0.3847	0.2200	0.5810
11	72	12	60,000	0.1218	0.0527	0.2567
12	72	12	80,000	0.9951	0.9605	0.9994
13	72	18	70,000	0.9785	0.8541	0.9972

Table 10.4 Input parameters for minimum acceptable Pr{No Leak} = P_{min}

Optimization					
Effect	Coeff	Coded level	Coeff * Effect	Decoded	Units
Intercept	0.132	1	0.132		NA
OD	1.5417	0.018024677	0.027788645	60.22	Inches
WT	2.1432	0.830706743	1.780370691	16.98	Inches
SMYS	3.9493	0.672261325	2.654961652	7,6722.61	Psi
		Logit:	4.595120988		
		exp(−logit):	0.010100999		
		Pr{Y = 1}:	**0.990000011**		
		Goal	**0.99**		
		Diff	**−1.12621E−08**		

$$SS_w = \sum_{i=1}^{n} w_i \left(x_i - \overline{X}_w\right)$$

$$se\left(\widehat{\beta}_0\right) = \sqrt{\frac{1}{\sum_{i=1}^{n} w_i} + \frac{\overline{X}_w^2}{SS_w}}$$

$$se\left(\widehat{\beta}_1\right) = \frac{1}{\sqrt{SS_w}}$$

$$cov\left(\widehat{\beta}_0, \widehat{\beta}_1\right) = -\frac{\overline{X}_w}{SS_w}$$

Back to the example. The optimal run was #9, which had OD at the center value (60 in.), and WT and SMYS at their high values (18 in., 80,000 psi, respectively). The highest predicted value, however, is achieved when all three factors are set to their highest levels, namely OD = 72, WT = 18, SMYS = 80,000. Under these conditions, the predicted probability of no leak is approximately 0.9994. The largest coefficient of the coded factors is for SMYS, so we can conclude that this is the most influential factor in determining the probability of no leak.

One might ask what values of the regressors would yield some minimally acceptable probability, call it P_{min}, of no leak. The EAS could use the MS Excel solver to compute the values of the regressors that yield at least a P_{min} probability of no leak. For example, suppose $P_{min} = 0.99$. Table 10.4 shows the computations and the solution as provided by the Excel solver function.

So, subject to the constraints that OD $\leq$ 72 in., WT $\leq$ 18 in., and SMYS $\leq$ 80,000 psi, the solution that yields an estimated $P_{min} = 0.99$ is OD $\approx$ 60 in., WT $\approx$ 17 in., SMYS $\approx$ 76,723 psi. Of course, it may be that such specifications would require custom manufacturing, so that it could be more economical to choose materials with specifications equal to those in run #9.

Example: A Discrete Factor

Logistic regression can be used to compare probabilities of success between discrete groups, much like ANOVA or t-tests are used to compare averages. We will consider a simple case of a single factor, call it X, with two discrete states. For convenience, the two states will be coded as $X = 1$ and $X = 0$. The response variable, Y, is of course binary, with its states coded as $Y = 1$ for "success" and $Y = 0$ for "non-success". As done earlier, we use the logit transformation, and assume it has a linear relationship to the factor, namely:

$$\lambda(X) = \ln(O) = ln\left(\frac{P(X)}{1 - P(X)}\right) = \beta_0 + \beta_1 X$$

$P(X)$ is the probability that $Y = 1$ given X.

The inverse logit is:

$$P(X) = \Pr\{Y = 1 | X\} = \frac{1}{1 + e^{-\lambda(X)}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

The way in which the coefficients are estimated is identical in nature to the case where the regressors were continuously-valued, namely via maximum likelihood. The only thing that differs is the interpretation.

The odds of $Y = 1$ given X are:

$$O(X) = \frac{P(X)}{1 - P(X)} = e^{\beta_0 + \beta_1 X}$$

The odds ratio is:

$$\frac{O(1)}{O(0)} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

Thus e^{β_1} is the number of times more likely to obtain a "success" when $X = 1$ than it is when $X = 0$. The natural logarithm of the odds ratio is called the "log odds ratio" and in the logistic case is simply β_1.

The value of the log odds ratio could be a design criterion. For example, an EAS may desire to improve a design by at least doubling its odds of achieving some goal. In that case, the EAS would want:

$$e^{\beta_1} = 2$$

Table 10.5 Summary of new vs. current App experimental results

App	Result	Count	Percent (%)
Current	Non-success	4	20.00
Current	Success	16	80.00
New	Non-success	2	10.00
New	Success	18	90.00

Or in other words:

$$\beta_1 = \ln(2) \approx 0.69315$$

Suppose a human factors engineer has tried to improve a computer "App" to increase the ease with which users can perform a task. The current App has an 80 % success rate for first-time users, so the odds of success are:

$$O_c = \frac{P_c}{1 - P_c} = \frac{0.8}{0.2} = 4$$

The EAS wished to increase the success rate to 90 %, so the odds of success using the new App would be:

$$O_n = \frac{P_n}{1 - P_n} = \frac{.9}{.1} = 9$$

The desired odds ratio is then:

$$\frac{O_n}{O_c} = \frac{P_n/(1 - P_n)}{P_c/(1 - P_c)} = \frac{9}{4} = 2.25$$

A new prototype App was created, and an experiment was performed with a sample of 40 subjects randomly selected from the target population. Half were randomly selected to use the current App, and half to use the new App. The results of the experiment are summarized in Table 10.5

Figure 10.5 shows the SAS Genmod code for analyses the data. The states for the App are coded as App = 0 for Current and App = 1 for New. Figure 10.6 shows the Genmod output. In this analysis, the "noint" option was used in the Proc Genmod. Genmod will then produce estimates of log odds for each App. The estimate for App = 0 is the log odds for the Current App, and the estimate for App = 1 is the log odds for the New App. Note that in the "App Least Squares Means" Table, the "Estimate" column is the log odds for each level of App. Thus, the estimated odds of success for the Current App are:

$$\widehat{O}_c \approx e^{1.3863} \approx 4.00$$

And the estimated odds of success for the New App are:

$$\widehat{O}_n \approx e^{2.1972} \approx 9.00$$

Thus it appears that the EAS has achieved his goal. Of course, the confidence intervals (see the columns labeled "Lower" and "Upper" in the App Least Squares Means Table) for the odds are fairly wide, due to the relatively small sample size.

```
libname stuff 'H:\Personal Data\Experimentation for Design &
Validation\Data & Analyses';

data calc;

  set stuff.d20141005_example_9_5_odds_ratio; /* do not include the
sas7bdat extension here */
/*************************************************************
*                                                            *
* variables: App = version of App (Current or New)           *
*            Subject = subject ID (1-20 for each App)        *
*            Result = binary response (1 = success, 0 = not) *
*                                                            *
*                                                            *
*                                                            *
*                                                            *
*************************************************************/

  run;

proc sort data=calc;
  by App;
  run;
proc means data=calc;
  var Result;
  by App;
  run;

proc genmod data=calc descending; * the "descending" parameter makes
the model for Pr{Y = 1} ;
  Class App;
  model Result = App/dist = bin link = logit noint;
  LSMEAN App/CL;
  output out=NLoutest PREDICTED=predprob LOWER=lcl UPPER=ucl;
  run;

data stuff.o20140921_odds_ratio_output;

  set NLoutest;

  run;
```

Fig. 10.5 SAS Genmod code—computer App example (single categorical factor)

Figure 10.7 shows the output from the JMP 11 "Fit Model" function. Note that this version of JMP does not allow the "no intercept" option for logistic regression. The estimate for the intercept minus the estimate for the "App" term is approximately −2.1972, which gives the log odds for the reciprocal of the odds for success using the New App. The point to consider is that interpreting the output of software, especially for logistic regression with discrete regressors must be done carefully.

For this simple example, the odds, and hence the log odds, for both Current and New Apps can be computed very easily. Recall that with the Current App, 16 out of 20, or 80 %, of the participants were successful. With the New App, 18 out of 20 were successful (90 %). The odds can be computed as:

```
                                The SAS System         08:23 Monday, October 6, 2014   1

------------------------------------------- App=0 -------------------------------------------

                                The MEANS Procedure

                            Analysis Variable : Result

 N            Mean         Std Dev         Minimum         Maximum
ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff
20       0.8000000       0.4103913               0       1.0000000
ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff


------------------------------------------- App=1 -------------------------------------------

                            Analysis Variable : Result

 N            Mean         Std Dev         Minimum         Maximum
ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff
20       0.9000000       0.3077935               0       1.0000000
ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff

                                The SAS System         08:23 Monday, October 6, 2014   2

                                The GENMOD Procedure

                                 Model Information

                            Data Set              WORK.CALC
                            Distribution           Binomial
                            Link Function             Logit
                            Dependent Variable       Result


                          Number of Observations Read          40
                          Number of Observations Used          40
                          Number of Events                     34
                          Number of Trials                     40


                                Class Level Information

                              Class      Levels    Values

                              App             2    0 1


                                   Response Profile

                             Ordered                  Total
                               Value    Result    Frequency

                                   1    1                34
                                   2    0                 6

PROC GENMOD is modeling the probability that Result='1'.


                                Parameter Information

                           Parameter       Effect       App

                           Prm1            Intercept
                           Prm2            App          0
                           Prm3            App          1
```

Fig. 10.6 SAS Genmod output—computer App example (single categorical factor)

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Log Likelihood		-16.5097	
Full Log Likelihood		-16.5097	
AIC (smaller is better)		37.0194	
AICC (smaller is better)		37.3437	
BIC (smaller is better)		40.3972	

The SAS System 08:23 Monday, October 6, 2014 3

The GENMOD Procedure

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		0	0.0000	0.0000	0.0000	0.0000	.	.
App	0	1	1.3863	0.5590	0.2906	2.4819	6.15	0.0131
App	1	1	2.1972	0.7454	0.7364	3.6581	8.69	0.0032
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Lagrange Multiplier Statistics

Parameter	Chi-Square	Pr > ChiSq
Intercept	.	.

App Least Squares Means

App	Estimate	Standard Error	z Value	Pr > |z|	Alpha	Lower	Upper
0	1.3863	0.5590	2.48	0.0131	0.05	0.2906	2.4819
1	2.1972	0.7454	2.95	0.0032	0.05	0.7364	3.6581

Fig. 10.6 (continued)

$$\widehat{O}_c = \frac{0.80}{0.20} = 4.00$$

and

$$\widehat{O}_n = \frac{0.90}{0.10} = 9.00$$

The standard error formula for logits when there is a single regressor simplifies even further when that regressor is discrete with only two states. The basic equations remain unchanged, however the following simplifications (Fleiss et al. 2003) can be substituted:

Ordinal Logistic Fit for Result

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	0.398656	1	0.797311	0.3719
Full	16.509708			
Reduced	16.908364			

RSquare (U)	0.0236
AICc	37.3437
BIC	40.3972
Observations (or Sum Wgts)	40

Measure	Training	Definition
Entropy RSquare	0.0236	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.0346	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.4127	∑ -Log(ρ[j])/n
RMSE	0.3536	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.2500	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.1500	∑ (ρ[j]≠ρMax)/n
N	40	n

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[non-success]	-1.7917595	0.4658475	14.79	0.0001*
App[Current]	0.40546511	0.4658475	0.76	0.3841

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
App	1	1	0.79731115	0.3719

Fig. 10.7 JMP output for logistic regression with a discrete regressor

$$\sum_{i=1}^{n} w_i = n_c p_c(1-p_c) + n_n p_n(1-p_n)$$

$$\overline{X}_w = \frac{n_n p_n(1-p_n)}{n_c p_c(1-p_c) + n_n p_n(1-p_n)}$$

$$SS_w = n_n p_n(1-p_n) - \frac{\{n_n p_n(1-p_n)\}^2}{n_c p_c(1-p_c) + n_n p_n(1-p_n)} = \frac{n_n p_n(1-p_n) n_c p_c(1-p_c)}{n_c p_c(1-p_c) + n_n p_n(1-p_n)}$$

Note that the subscripts "c" and "n" refer to Current or New App, respectively.

Key Points

- Odds, the ratio of the probability of "success" and its compliment, are a measure of the chance of success.
- The log of odds gives rise to the logit transformation.
- Binary logistic regression is a means of creating a predictive model for probabilities when the response variable is binary.
- The parameters in a binary logistic regression model are the log odds.

Exercises and Questions

1. Suppose equal numbers of two varieties of rice, A and B, are planted in a greenhouse/lab. After a fixed [period, the number of shoots that sprouted were counted. For variety A, 90 % of the seeds sprouted, and 85 % of the variety B seeds sprouted. What are the odds for sprouting for each, and the odds ratio of variety A to variety B?
2. Assuming that for both varieties in question 1, $n = 100$, compute 95 % confidence intervals for the logits of the probability of sprouting for each variety. Then transform the confidence interval limits into limits on the probability of sprouting (inverse logit, or logistic transform).

Chapter 11
Reliability, Life Testing, and Shelf Life

The Reliability and Related Functions

Everything will fail, eventually. Reliability is a probability that a system (or a component) will fail no sooner than t time units from the time it begins operating. Reliability has many manifestations, or one might consider reliability as a special case of probabilities that some specific type of event will occur no sooner than t units from some initial reference time. Other manifestations include survival of patients having a particular disease, or the shelf life of drugs. For simplicity, we will refer to reliability in terms of time to failure, with the understanding that this time could be the time from a reference point to the occurrence of some specific type of event (such as death, progression of disease, a drug concentration in a human body drops below some threshold, or potency/reactivity loss). The point is to derive a model by which the probability of interest can be predicted, and to incorporate design parameters into this model. In this way, we hope to help the EAS design a system so that it will have a desired reliability for a particular time-to-event.

To begin, we borrow from elementary chemical kinetics (Whitten et al. 2004), and consider a first-order system. Suppose we can measure a response variable that indicates the degree to which a system is operating properly (we realize this could be challenging in many cases, but it is our point of departure.) Let $Y(t)$ be the response variable observed at time t. A first-order differential equation model that describes the change in this response over time can be expressed as:

$$\frac{dY(t)}{dt} = -\lambda Y(t)$$

With initial condition $Y(0) = y_{\mathbf{0}}$, the solution to the equation is:

$$Y(t) = y_0 e^{-\lambda t}$$

Some may recognize this as the equation governing first-order chemical reaction kinetics (Whitten et al. 2004), or the equation for radioactive decay (Rutherford 1900). The parameter $\boldsymbol{\lambda}$ is called the rate parameter (as in the rate of reaction). If we presume that the response variable $Y(t)$ is a decreasing function of time, as in the case of analyte concentrations in chemical reactions, then it is a maximum at $t = 0$, and its maximum value is y_0. Thus we can express the response as a proportion of its initial value, namely:

S.A. Pardo, *Empirical Modeling and Data Analysis for Engineers and Applied Scientists*,
DOI 10.1007/978-3-319-32768-6_11

$$\frac{Y(t)}{y_0} = e^{-\lambda t}$$

The proportion can be thought of at a probability, namely the probability that the value of the response is not zero after t time units. So, if we replace $Y(t)$ with a new variable, namely $T =$ the time at which $Y(t)$ is zero, then we can express the value of $Y(t)$ as a proportion of its initial value with $\Pr\{T \geq t\}$, or:

$$Pr\{T \geq t\} = e^{-\lambda t}$$

It turns out that this is the complement of the cumulative distribution function of an exponential random variable. That is,

$$\Pr\{T \leq t\} = 1 - \Pr\{T \geq t\} = 1 - e^{-\lambda t}$$

In the language of Reliability, the parameter λ is called the failure rate. The exponential time to failure variable is characterized by a constant failure rate. That is, potentially the rate of failure could change with time, t. if $h(t)$ represents the failure rate, then for the exponential time to failure variable,

$$h(t) = \lambda \ \forall t$$

The exponential time to failure variable is related to a discrete random variable with a Poisson distribution. The variable, X, is the number of failures (or events) within a fixed length of time. It has a Poisson distribution if:

$$Pr\{X = x|\lambda, t\} = \frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

The value of t is a fixed length of time, and λ is the (constant) failure rate.

Suppose that the failure rate actually changes with time. For example, the rate of failure could be fairly high initially, then drop after a "burn-in" period to a constant, and then climb back up after the product reaches a "wear-out" time. Such a failure rate function is sometimes called a "bath-tub" curve, as illustrated in Fig. 11.1.

The failure rate function is also called the hazard rate function.

To generalize our initial first-order differential equation model for the response variable $Y(t)$, we could replace the constant λ with $h(t)$:

$$\frac{dY(t)}{dt} = -h(t)Y(t)$$

Again using the initial condition $Y(0) = y_0$, the solution is

$$Y(t) = y_0 e^{-\int_0^t h(\tau)d\tau}$$

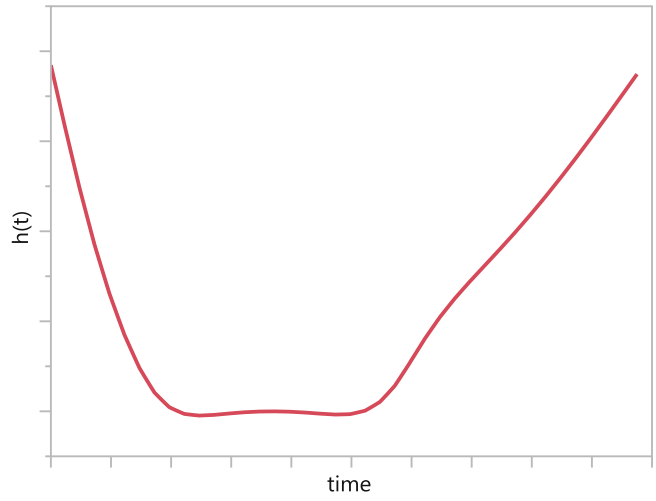


Fig. 11.1 "Bath-tub" failure rate curve

Of course, if $h(t) = \lambda$ (a constant), then:

$$\int_0^t h(\tau)d\tau = \lambda t$$

The function:

$$H(t) = \int_0^t h(\tau)d\tau$$

is called the cumulative failure rate or cumulative hazard rate function for our new random variable T, time to failure. So, in general, using our slightly generalized first order kinetics model,

$$Pr\{T \geq t\} = e^{-\int_0^t h(\tau)d\tau} = e^{-H(t)}$$

If we define the reliability function to be:

$$R(t) = Pr\{T \geq t\} = e^{-\int_0^t h(\tau)d\tau} = e^{-H(t)}$$

Then the reliability function can be thought of as a curve in time. We also have some potentially useful relationships:

$$-lnR(t) = H(t)$$

$$h(t) = \frac{dH(t)}{dt}$$

Another related pair of functions is:

$$F(t) = 1 - R(t) = 1 - e^{-H(t)}$$

$$G(t) = -lnF(t)$$

$F(t)$ is the cumulative distribution function for time-to-failure.

Thus we have the relationship:

$$R(t) = 1 - e^{-G(t)}$$

The design problem is to state either the reliability function or cumulative failure rate function in terms of design parameters, and then choose values of those parameters that yield the desired reliability curve.

Obtaining an Empirical Reliability Model

Suppose the EAS does not know the specific form of $R(t)$, $H(t)$, or $h(t)$. She or he could perform an experiment to obtain a polynomial approximation. We will call the experiment a Life Test. In this experiment, n items (devices, systems, components) will be "started" and allowed to run until failure. The elapsed time from start to failure will be recorded. Suppose the n times to failure are ordered form shortest to longest. Call these times $t_1, t_2, \ldots, t_k, \ldots, t_n$. These times are referred to as "order statistics" (Conover 1999). Compute the empirical reliability function:

$$\widehat{R}(t_k) = 1 - \frac{k}{n} = \frac{n-k}{n} \qquad k = 1,\ n$$

or the empirical cumulative distribution function:

$$\widehat{F}(t_k) = 1 - \widehat{R}(t_k) = \frac{k}{n} \qquad k = 1, n$$

For a lack of a better term, we will call this formula, or estimator, the empirical maximum likelihood (EML) estimator. Now suppose that the cumulative hazard rate function, $H(t)$, can be approximated by a low order polynomial in t, for example:

$$H(t) = \beta_0 + \beta_1 t + \beta_2 t^2$$

Given the relation:

$$-lnR(t) = H(t)$$

It may be advantageous to approximate $G(t)$ as a second-order polynomial, i.e.:

$$G(t) = \beta_0 + \beta_1 t + \beta_2 t^2$$

The EAS can now obtain via least squares an estimate of the parameters β_k, and thus an approximation formula for $R(t)$. That is, the estimated approximation formula would be:

$$\tilde{R}(t) = 1 - \exp(-G(t)) = 1 - \exp\left(-\widehat{\beta}_0 - \widehat{\beta}_1 t - \widehat{\beta}_2 t^2\right)$$

The $\widehat{\beta}_i$ are the least squares estimates of the $H(t)$ approximation formula parameters. The approximation formula could be used to interpolate values of $R(t)$, but interpolation is not its most important use. More importantly, it can be used as a design tool.

Relating the β_i to Design Parameters

Consider a system with design factors or parameters, $x_1, x_2, \ldots, x_j, \ldots, x_m$. The EAS could perform a designed experiment in these factors, where the response is T = time to failure. At each run in the experiment, the reliability function approximation can be fit to the data. In this way, the reliability parameters can be related to the design factors. That is, we want to obtain an approximation formula that allows us to predict each of the parameters, β_0, β_1, and β_2, as polynomial functions of the design factors $x_1, x_2, \ldots, x_j, \ldots, x_m$. These approximating polynomials are obtained once again via least squares. Pardo (2009) describes how this methodology can be used for designing solid dosage forms for pharmaceuticals.

An example will help. Suppose the EAS wants to design an artificial hip joint replacement. The design will be a metal-on-polyethylene system. The questions are which metal material (steel or aluminum), the shape of the polyethylene (PE) pin (or truncated cone), and the type of lubricant (albumin or gamma globulin) to use. Normally, hip joint replacements are intended to last 15–20 years with a high degree of probability. That is, the expectation is that the reliability at 15 years should be at least 90 %. The EAS decides to perform a 2^3 factorial experiment in the three factors, metal, PE, and lube. Table 11.1 shows the eight different prototypes to be constructed.

A simulator device was constructed, so that years of wear could be achieved without actually implanting the devices in people and waiting for them to fail. A total of n = 10 units of each prototype were constructed and tested on the simulator until they failed.

Table 11.2 shows the time-to-failure data from the simulator. For each prototype, the times have been sorted from shortest to longest, and the empirical reliability has been calculated.

For each of the prototypes, a second order polynomial approximation was fit to

$$\widehat{G}(t) = -ln\widehat{F}(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \epsilon$$

Table 11.3 shows the resulting estimates of the parameters. In addition, the predicted reliability at t = 20 years is computed for each prototype.

Table 11.1 2^3 factorial experiment in hip replacement prototypes

Prototype	Metal	PE pin	Lube
1	Steel	Cone	Albumin
2	Steel	Cone	GG
3	Steel	Cylinder	Albumin
4	Steel	Cylinder	GG
5	Aluminum	Cone	Albumin
6	Aluminum	Cone	GG
7	Aluminum	Cylinder	Albumin
8	Aluminum	Cylinder	GG

Table 11.2 Time-to-failure (years) data from simulator experiments

Rhat	P1	P2	P3	P4	P5	PS	P7	PB
0.9	16.23	18.37	16.36	14.30	21.48	24.05	19.71	22.28
0.8	18.25	18.56	18.51	16.84	22.01	24.63	20.93	22.57
0.7	18.95	18.72	18.72	17.76	22.90	26.86	21.43	22.95
0.6	19.27	19.21	18.95	19.86	23.20	27.35	21.67	24.17
0.5	19.41	20.94	18.97	20.38	23.86	29.09	23.53	25.29
0.4	20.05	21.83	19.59	20.63	24.08	29.51	24.07	25.66
0.3	20.42	22.01	19.79	20.77	24.71	30.10	24.57	25.66
0.2	20.91	22.62	20.37	21.99	25.57	30.50	24.93	26.12
0.1	21.19	22.80	20.71	23.57	25.59	31.00	25.40	26.53
0	21.95	24.05	21.21	24.10	26.64	31.51	25.56	27.63

Table 11.3 Parameter estimates for second-order polynomial fits to $H(t)$

	$G(t) = B0 + B1*t + B2*t^2$						
Prototype	Metal	PE pin	Lube	B0	B1	B2	R(20)
P1	Steel	Cone	Albumin	15.11	−1.04	0.02	0.459
P2	Steel	Cone	GG	28.92	−2.38	0.05	0.610
P3	Steel	Cylinder	Albumin	16.06	−1.08	0.02	0.351
P4	Steel	Cylinder	GG	9.65	−0.67	0.01	0.506
P5	Aluminum	Cone	Albumin	58.80	−4.43	0.08	0.973
P6	Aluminum	Cone	GG	16.97	−0.88	0.01	0.975
P7	Aluminum	Cylinder	Albumin	29.81	−2.21	0.04	0.869
P8	Aluminum	Cylinder	GG	41.50	−2.92	0.05	0.974

If the design objective is to achieve a 95 % reliability at $t = 20$ years, it is clear that Steel is not adequate. What is not clear is whether there is an interaction between The PE Pin and the Lube. It appears that the Cone shape may be superior to Cylinder, but it may be that Cylinder with GG is at least as good as Cone with Albumin.

An ANOVA could be performed to determine which factors have an effect on time-to-failure, at least on the average. Figure 11.2 shows the output from JMP.

The residual by predicted value plot indicates that the noise variance is constant over the range of failure times, so the p-values are probably valid. The only statistically significant (i.e., probably repeatable) interaction effect was between Metal and Lube. However, the difference in average failure time for Cone, between Albumin and GG (~2.9 years), and the difference for Cylinder between Albumin and GG (~2.2 years) were not much different that the difference for Aluminum between Albumin and GG (~3.1 years) and for Steel between Albumin and GG (~1.0 years). The greater discrepancy was for the Metal/Lube interaction, which was apparently large enough to make the interaction significant.

Table 11.4 shows the mean and standard deviations (SD) for time-to-failure and the mean predicted reliability at 20 years for each level of the three factors. The maxima for both mean failure time and reliability are shown in bold font. Note that the conditions that yield maximal values are Aluminum, Cone, and GG (P6). The average time-to-failure for this condition is 28.46 years, and the predicted reliability at 20 years is 0.975. A close second place is Aluminum, Cylinder, GG (P8), with a predicted reliability of 0.974. A very close third place is Aluminum, Cone, Albumin (P5), with predicted $R(20) = 0.973$. The lower SD for P5 indicates that it may yield more consistent wear and hence more predictable time-to-failure. The lower limit of the 95 % confidence interval for predicted reliability at 20 years for P5 is 0.9585, which actually exceeds our design criterion (0.950).

The ANOVA has indicated several things we might not have known without this experiment:

1. Metal type, Pin type, and Lube type all significantly affect the time-to-failure;

Response Time

Summary of Fit

RSquare	0.68492
RSquare Adj	0.659023
Root Mean Square Error	2.10263
Mean of Response	22.55525
Observations (or Sum Wgts)	80

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	701.5640	116.927	26.4479
Error	73	322.7368	4.421	**Prob > F**
C. Total	79	1024.3008		<.0001*

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Metal	1	1	531.37741	120.1925	<.0001*
PE Pin	1	1	39.67744	8.9747	0.0037*
Lube	1	1	82.25568	18.6055	<.0001*
Metal*PE Pin	1	1	12.49781	2.8269	0.0970
Metal*Lube	1	1	22.17618	5.0160	0.0282*
PE Pin*Lube	1	1	13.57952	3.0716	0.0839

Residual by Predicted Plot

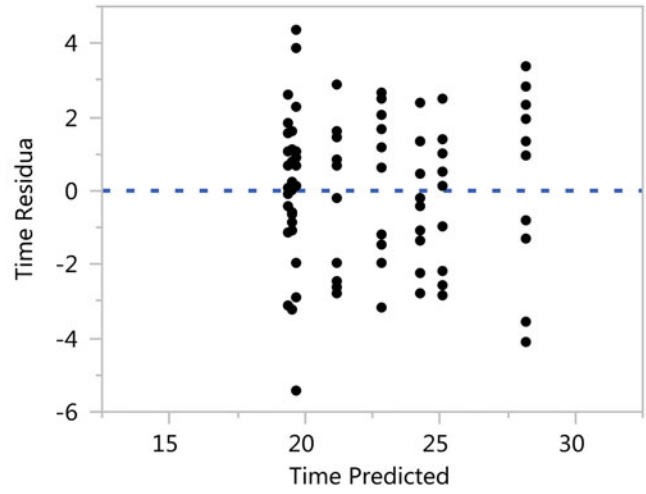


Fig. 11.2 ANOVA for hip replacement time-to-failure

Effect Details

Metal

Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
Aluminum	25.132500	0.33245494	25.1325
Steel	19.978000	0.33245494	19.9780

PE Pin

Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
Cone	23.259500	0.33245494	23.2595
Cylinder	21.851000	0.33245494	21.8510

Lube

Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
Albumin	21.541250	0.33245494	21.5413
GG	23.569250	0.33245494	23.5693

Metal*PE Pin

Least Squares Means Table

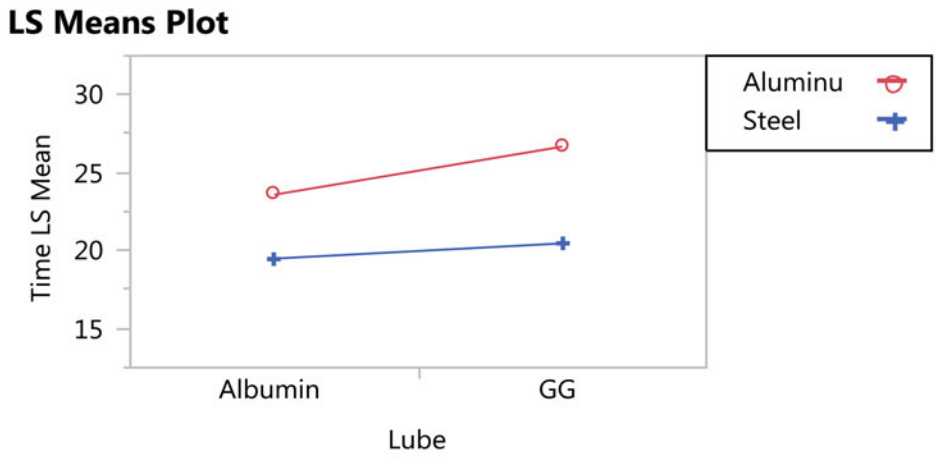
Level	Least Sq Mean	Std Error
Aluminum,Cone	26.232000	0.47016229
Aluminum,Cylinder	24.033000	0.47016229
Steel,Cone	20.287000	0.47016229
Steel,Cylinder	19.669000	0.47016229

Metal*Lube

Least Squares Means Table

Level	Least Sq Mean	Std Error
Aluminum,Albumin	23.592000	0.47016229
Aluminum,GG	26.673000	0.47016229
Steel,Albumin	19.490500	0.47016229
Steel,GG	20.465500	0.47016229

Fig. 11.2 (continued)



PE Pin*Lube

Least Squares Means Table

Level	Least Sq Mean	Std Error
Cone,Albumin	21.833500	0.47016229
Cone,GG	24.685500	0.47016229
Cylinder,Albumin	21.249000	0.47016229
Cylinder,GG	22.453000	0.47016229

Fig. 11.2 (continued)

Table 11.4 Mean time-to-failure and predicted R(20)

			Time-to-failure						
Prototype	Metal	PE pin	Lube	N	Mean	SD	Pred. R(20)	LCL-R(20)	UCL-R(20)
P5	**Aluminum**	**Cone**	**Albumin**	**10**	**24.00**	**1.66**	**0.973**	**0.9585**	**0.9826**
P6	**Aluminum**	**Cone**	**GG**	**10**	**28.46**	**2.62**	**0.975**	**0.9117**	**0.9930**
P7	Aluminum	Cylinder	Albumin	10	23.18	2.08	0.869	0.8290	0.9001
P8	Aluminum	Cylinder	GG	10	24.89	1.81	0.974	0.9039	0.9930
P1	Steel	Cone	Albumin	10	19.66	1.64	0.459	0.3844	0.5239
P2	Steel	Cone	GG	10	20.91	2.06	0.610	0.4510	0.7229
P3	Steel	Cylinder	Albumin	10	19.32	1.37	0.351	0.2158	0.4627
P4	Steel	Cylinder	GG	10	20.02	3.02	0.506	0.4434	0.5610

2. Although related, the probability that time-to-failure exceeds 20 years and the mean time to failure are not surrogates for each other;
3. The effect of Lube may in fact depend on the type of Pin; The interaction effect may be more apparent in terms of consistency of wear;

Had we not constructed the low-order polynomial approximations for $G(t)$, we would not have been able to generate predictions for reliability at 20 years. Had we not used a designed experiment, we might not have realized that all three factors actually do affect time-to-failure. Furthermore, without both the designed experiment and the models for $G(t)$, we might not have discovered that the Pin type and Lube type interact with each other in terms of reliability, even though they do not appear to interact in terms of mean time-to-failure.

Censored Time-to-Failure

Sometimes limits are placed on the length of time over which experimental units will be observed, or on the number of units (out of a sample of size n) that will be allowed to fail before the test is stopped. Such restrictions are called censoring. Stopping the test at a fixed time is referred to as Type I censoring (Mann et al. 1974), and stopping the test after a fixed number of units fail is called Type II censoring (Mann et al. 1974) The question is how to estimate reliability in the face of such censoring. We will address Type I censoring first.

Suppose the EAS did a life test with a sample of n experimental units, and stopped the test after T_{max} time units. Out of the n units on test, only $r < n$ actually failed. The remaining $n - r$ units were still functioning at time T_{max}. Suppose further that the r failure times are sequenced from shortest to longest, and that $t_1, t_2, t_3, \ldots, t_r$ represent those order statistics. So the empirical reliability function could be represented as it was when there was no censoring:

$$\widehat{R}(t_k) = 1 - \frac{k}{n} = \frac{n-k}{n} \qquad k = 1,\ r$$

The is the EML with right-censoring (EMLC). The EAS could treat this computed empirical reliability function in the exact same way he or she did when there was no censoring. There is another method for computing an empirical reliability function at times other than those which were explicitly observed failure times. The method is due to Kaplan and Meier (1958), and is described in Lee (1992). The formula for the Kaplan–Meier (K–M) estimator is:

$$\widehat{R}(t) = \prod_{t_{k \le t}} \frac{n-k}{n-k+1}$$

where t_k are the failure times for uncensored observations. At $t = t_k$, this formula can be written as:

$$\widehat{R}(t_k) = \widehat{R}(t_{k-1}) \frac{n-k}{n-k+1}$$

The K–M estimator is particularly useful if censoring can occur even though the time has not exceeded T_{max}. For example, a unit may fail during the test, but due to some cause other than the particular failure mode of interest.

The variance, and thus standard error, of the K–M reliability estimate, can be approximated (Lee 1992) by:

$$V\left(\widehat{R}(t_k)\right) \approx \widehat{R}^2(t_k) \sum_{i=1}^{k} \frac{1}{(n-i)(n-i+1)}$$

The approximate standard error is the square root of this variance approximation

Using either the EML, EMLC, or K–M estimators, the highest reliability estimate is at failure time t_1, and it is $(n - 1) / n$. The lowest reliability value would be at tr. If $r = n$, then EML = EMLC, and at t_r, $R(t_r) = 0$. For the K–M estimator, this is not the case if there are any intermediate censored failures (i.e., if a unit fails before T_{max} in some mode other than the one(s) of interest).

As an example, consider an EAS testing a circuit board. She is concerned about a particular component failing. She puts $n = 15$ boards on test, and sets $T_{max} = 1000$ h. Some of the boards fail before T_{max}, and some do not. Of the boards that fail, some of them had a failure in a component other

Table 11.5 Censored failure times with Kaplan–Meier estimates

k	Time-to-failure	Censored?	(n − k)/(n − k + 1)	$R(t_k)$ K–M	Factor	SE	LCL	UCL
1	870	Uncensored	0.9333	0.9333	0.00476	0.06441	0.8071	1.0000
2	872	Uncensored	0.9286	0.8667	0.01026	0.08777	0.6946	1.0000
3	884	Uncensored	0.9231	0.8000	0.01667	0.10328	0.5976	1.0000
4	889	Uncensored	0.9167	0.7333	0.02424	0.11418	0.5095	0.9571
5	909	Uncensored	0.9091	0.6667	0.03333	0.12172	0.4281	0.9052
6	915	Uncensored	0.9000	0.6000	0.04444	0.12649	0.3521	0.8479
7	916	Uncensored	0.8889	0.5333	0.05833	0.12881	0.2809	0.7858
8	932	Uncensored	0.8750	0.4667	0.07619	0.12881	0.2142	0.7191
9	937	Censored						
10	939	Uncensored	0.8333	0.3889	0.10952	0.12870	0.1366	0.6411
11	951	Uncensored	0.8000	0.3111	0.15952	0.12426	0.0676	0.5547
12	962	Uncensored	0.7500	0.2333	0.24286	0.11499	0.0080	0.4587
13	979	Uncensored	0.6667	0.1556	0.40952	0.09955	0.0000	0.3507
14	1000	Censored						
15	1000	Censored						

than the one of interest. Those boards that failed due to other components are censored. Table 11.5 shows the failure times, whether the failures were censored or not, the intermediate calculation $(n - k) / (n - k + 1)$, the K–M estimator for $R(t_k)$, and an approximate 95 % confidence interval (LCL, UCL) for the reliability at each uncensored failure time. The column labeled "factor" is the summation term:

$$\sum_{i=1}^{k} \frac{1}{(n-i)(n-i+1)}$$

Note that for censored observations, no computation is made. The ninth failure time was censored, as this unit failed in a mode that was not of interest to the EAS.

Regardless of how the reliability estimates were obtained (EML, EMLC, K–M), the computations for $\widehat{F}(t_k)$, the polynomial approximation of $G(t)$, and the computation of predicted values for $R(t)$ follow the same procedures.

Accelerated Life Tests

In the case of the hip replacement problem, the EAS had a simulator that could simulate years of life in a fairly short time. Sometimes, the simulation of life is performed by subjecting experimental units to some condition which is presumed to accelerate the failure process in such a way as to allow the EAS to predict the increase in the failure rate. Commonly, temperature is used as the accelerating condition. Presuming that the increase in failure rate is proportional to an increase in the rate of a chemical reaction, a model called the Arrhenius Reaction Rate Law (Mann et al. 1974) is employed to relate the failure rate to temperature. The Arrhenius model is:

$$\lambda_P = A exp\left(\frac{-E/K}{P}\right)$$

The constant A is specific to the particular materials and reactions that underlie the failure mode. The constant E is called the energy of activation, and is also specific to the materials and chemical reactions involved in failure. The letter K stands for Boltzmann's constant, and P is the temperature in degrees Kelvin (we use the letter P, for "parameter", so as to not confuse it with T for time-to-failure). So λ_P is the average failure rate at temperature P. The usual presumption is that the time-to-failure follows an exponential distribution, so that at temperature P, the reliability function is given by:

$$R\left(t\middle|P\right) = e^{-\lambda_P t}$$

The question is how to determine the degree of acceleration achieved by exposing experimental units to a particular temperature, say P_A. The first problem is to estimate the parameters A and $B = -E/K$ (note that B is just a normalized energy of activation). The answer depends on the nature of the data. Life tests may be performed by placing n units into a temperature chamber, at a fixed temperature, P_A, for a given time, T. At time T, the units are taken out and inspected or tested, and the number of units that "survive", $S = s$, are counted. Assuming that the time-to-failure is exponentially distributed, and an estimate of the reliability at time T is $p_A = \frac{s}{n}$, then the reliability is given by:

$$\widehat{R}\left(T\middle|P_A\right) = e^{-\lambda_A T} = p_A = \frac{s}{n}$$

Solving for λ_A gives an estimate for the failure rate:

$$\widehat{\lambda}_A = \frac{-\ln(p_A)}{T} = \frac{-\ln\left(\frac{s}{n}\right)}{T}$$

Suppose an experiment was performed where n_1 units were put on test for T time units at temperature P_1 and another n_2 units put on test for T time units at temperature P_2. Then we would have two equations in the two unknowns A and B:

$$-ln\left(\frac{s_1}{n_1}\right) = TAexp\left(\frac{-B}{P_1}\right)$$

$$-ln\left(\frac{s_2}{n_2}\right) = TAexp\left(\frac{-B}{P_2}\right)$$

Taking logs on both sides yields:

$$\ln\left(-ln\left(\frac{s_1}{n_1}\right)\right) = lnT + lnA - \frac{B}{P_1}$$

$$\ln\left(-ln\left(\frac{s_2}{n_2}\right)\right) = lnT + lnA - \frac{B}{P_2}$$

These in turn yield the solutions for the estimates:

$$\widehat{B} = P_1\left[ln\widehat{A} - ln\left(-ln\frac{s_1}{n_1}\right) + lnT\right]$$

$$ln\widehat{A} = \frac{P_2}{P_2 - P_1}\left[ln\left(-ln\frac{s_2}{n_2}\right) + \frac{P_1}{P_2}ln\left(-ln\frac{s_1}{n_1}\right) + \frac{P_1}{P_2}lnT\right]$$

Thus, an expression for the estimate of parameter B in only known quantities is:

$$\widehat{B} = P_1\left[\frac{P_2}{P_2 - P_1}\left[ln\left(-ln\frac{s_2}{n_2}\right) + \frac{P_1}{P_2}ln\left(-ln\frac{s_1}{n_1}\right) + \frac{P_1}{P_2}lnT\right] - ln\left(-ln\frac{s_1}{n_1}\right) + lnT\right]$$

These estimates will allow the EAS to determine how much acceleration was achieved at any given temperature, P_2, compared to a lower temperature, P_1. The estimates of B and A may be useful for future experiments or tests, provided that the materials involved in such tests are at least similar if not identical to those used to obtain the estimates.

Suppose that the EAS has at least an hypothetical failure rate desired at say 25 °C = 25 + 273 = 298°K = P_0. Such a failure rate might be determined by having a specification or requirement that the reliability at time T_0 must be at least r_0. Assuming the exponential time-to-failure, the failure rate is given by:

$$\lambda_0 = \frac{-\ln(r_0)}{T_0}$$

Now suppose that the EAS wants to accelerate the failure process k times, so that the actual test time would need to be $T_a = \frac{T_0}{k}$ ("a" stands for "accelerated"). This would also mean that the failure rate under the accelerated conditions would need to be:

$$\lambda_a = k\lambda_0$$

The EAS must choose a temperature, $P_a > P_0$, to achieve the desired acceleration.

Using the Arrhenius equation:

$$k = \frac{\exp(-B/P_a)}{\exp(-B/P_0)} = exp\left(B\left(\frac{1}{P_0} - \frac{1}{P_a}\right)\right)$$

Since k is actually given (i.e., the desired acceleration to allow the test to occur in a short enough time), and P_0 is known, the equation can be solved for P_a:

$$P_a = \frac{BP_0}{B - P_0 lnk}$$

The only thing required is a value for B, the normalized energy of activation. A simple experiment as described earlier can be used to obtain an estimate of B.

Recall that test temperatures should be expressed in degrees Kelvin ($°K$) when using these equations.

There are two potential drawbacks to the procedures described for determining parameters of an accelerated life test. First, we have assumed that the time-to-failure is affected by temperature in the way described by the Arrhenius equation. Secondly, we have assumed that time-to-failure has an exponential distribution. Both of these assumptions stem from a more fundamental assumption that the failure process is related to a first-order chemical reaction (Chow 2007). While these assumptions may not be completely valid, they may provide at least a practical approach to determining the amount of acceleration achieved by putting units on test at temperature P_a for time T_0.

As an example, consider a life test with $P_0 = 25\ ^\circ\text{C} = (25 + 273 = 298)^\circ\text{K}$ and P_a is $40\ ^\circ\text{C} = (40 + 273 = 313)^\circ\text{K}$. At P_0, a test with $n_0 = 30$ units is performed for $T_0 = 720$ h. At P_a, the test is also performed for $T_0 = 720$ h. with $n_2 = 30$ units. At P_0, the number of "surviving" units was $s_0 = 29$. At P_a, the number of operating units after 720 h. was $s_a = 20$. Figure 11.3 shows some R code, together with session window output, for computing the estimates of A, B, and k.

Note that due to the vectorized nature of the computations in R, and the fact that the data were entered with separate rows for the 25 and 40 °C inputs, R computes the parameter estimates as vectors of length 2.

In this example, increasing temperature from 25 to 40 °C results in an acceleration of about 18.8 times. Thus, the 720 h at 40 °C is equivalent to 18.8*720 = 13,536 h at 25 °C, or about 1.54 years. If the EAS wished to simulate 2 years = 17,532 h with a test of 720 h, then he or she would need a test temperature that would yield an acceleration of approximately 17,532/720 = k = 24.35 times. Using the equation for Pa, given B and k, yields:

$$P_a = \frac{BP_0}{B - P_0 lnk} = \frac{18231.39 * 298}{18231.39 - 298 * \ln(24.35)} \approx 314.1^\circ\text{K} = 41.1^\circ\text{C}$$

So, it turns out that a relatively small change (1.1 °C) in the test temperature would give the desired acceleration.

Finally, the EAS may want to have an estimate of the failure rate at the new test temperature of 41.1 °C. Using the Arrhenius equation:

$\lambda_P = Aexp\left(\frac{-B}{P}\right) \approx 0.0012$, or approximately 0.0012 failures per hour at 41.1 °C. The estimated failure rate at 25 °C would then be:

$$\lambda_0 = \frac{\lambda_a}{k} \approx \frac{0.0012}{24.35} \approx 0.000049$$

The estimated probability that the device would last 2 years is given by the exponential reliability function:

$$R\left(2 \text{ years} \middle| 25^\circ\text{C}\right) = e^{-0.000049*17532} \approx 0.4236,$$

or about a 42.36 % chance that the device will not fail before 2 years. Recall that at 25 °C, there were 29 out of 30 parts that survived after only 720 h. It does not seem unreasonable to expect such a low reliability at 2 years. The question of why may require a designed experiment in the features the EAS determines may have a critical impact on reliability. Armed with the Arrhenius parameter estimates, the EAS may perform the experiment at 41.1 °C for 720 h per each experimental set of prototypes. If the only information per unit tested is whether at 720 h it was or was not operational, logistic regression methods may well be appropriate to determine which device features are most critical, and what particular choices for each critical feature may be optimal.

Stability and Shelf Life

Shelf life is the amount of time an item may remain unused and still perform adequately. Chemical products, such as pharmaceuticals, paint, solvents, and even food items are subject to the problem of estimating, controlling, and elongating shelf life. Shelf life is closely related to reliability, although the two are not identical concepts. We will treat the case where there is a continuously-valued quality

```
setwd("H:\\Personal Data\\Experimentation for Design & Validation\\Data & Analyses\\")
df1 <- read.csv("20141110 accelerated life test.csv")
attach(df1)
# inputs:
# Test = index
# TempC = test temp in degrees C
# Time = test time in hours
# n = number of units on test
# s = number surviving units after test
#
# the R function "log" is the natural logarithm
#
TK1 <- TempC[1] + 273
TK2 <- TempC[2] + 273
n1 <- n[1]
n2 <- n[2]
s1 <- s[1]
s2 <- s[2]
lnA <- (TK2)/(TK2 - TK1)*(log(-log(s2/n2)) + TK1*log(-log(s1/n1))/TK2 + log(Time))
B <- TK1*(lnA - log(-log(s1/n1)) + log(Time))
K <- exp(-B/TK2) / exp(-B/TK1)
df2 <- cbind(lnA,B,K)
write.csv(df2,file="20141110.acceleration.parameters.csv")

> setwd("H:\\Personal Data\\Experimentation for Design & Validation\\Data &
Analyses\\")
> df1 <- read.csv("20141110 accelerated life test.csv")
>
> attach(df1)
>
>
> # inputs:
> # Test = index
> # TempC = test temp in oC
> # Time = test time in hours
> # n = number of units on test
> # s = number surviving units after test
> #
> # the R function "log" is the natural logarithm
> #
> TK1 <- TempC[1] + 273
> TK2 <- TempC[2] + 273
> n1 <- n[1]
> n2 <- n[2]
> s1 <- s[1]
```

Fig. 11.3 R code for accelerated life test parameter estimation

```
> s2 <- s[2]
>
> lnA <- (TK2)/(TK2 - TK1)*(log(-log(s2/n2)) + TK1*log(-log(s1/n1))/TK2 +
log(Time))
> B <- TK1*(lnA - log(-log(s1/n1)) + log(Time))
> K <- exp(-B/TK2) / exp(-B/TK1)
>
> df2 <- cbind(lnA,B,K)
>
> write.csv(df2,file="20141110.acceleration.parameters.csv")
> lnA
[1] 51.21562 51.21562
> B
[1] 18231.39 18231.39
> K
[1] 18.76342 18.76342
>
```

Fig. 11.3 (continued)

response variable. Furthermore, we will assume that this variable is monotonic with respect to time spent "on the shelf", and that it is continuously degrading. As in the cases of Chaps. 5–8, the shelf-life variable can be optimized over the list of critical features, components, or factors. We will concentrate on how to estimate shelf life, following methods described by Chow (2007).

Assume that the quality variable, Y, is linearly related to shelf time, S, i.e.:

$$Y = \beta_0 + \beta_1 S + \varepsilon$$

There are two sorts of shelf-life problems; (1) find the time, S_e, such that there is a $100p$ % chance that the value of Y will still be acceptable; (2) determine the probability, p, that Y is acceptable at a pre-determined "warranty" time, S_w. For both of these problems, a prediction interval approach will be employed.

Recall that earlier, the standard error of a predicted value from a polynomial regression model was presented. As a special case of a linear model in a single regressor, namely time, the formula for a future predicted value at time $= s_k$ is:

$$SE\left(\widehat{Y}\,|s_k\right) = \sigma\sqrt{1 + \frac{1}{n} + \frac{(s_k - \overline{s})^2}{\sum_{i=1}^{n}(s_i - \overline{s})^2}}$$

The standard deviation of the noise variable, ε, namely σ, is estimated by the root mean square error of the regression fit, $\widehat{\sigma}$. The lower confidence limit for a predicted value of Y at time $= s_k$ is then:

$$\widehat{Y}_L = \widehat{Y}_{s_k} - t_{1-\alpha}(n-2)\widehat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(s_k - \overline{s})^2}{\sum_{i=1}^{n}(s_i - \overline{s})^2}}$$

To solve problem 1, we will employ the inverse regression approach (Draper and Smith 1998). To find the time, s_k, at which it is expected the response variable, Y, would be no lower than the lower specification limit, L, with probability $1 - \alpha$, solve the equation for s_k:

$$L = \widehat{Y}_{s_k} - t_{1-\alpha}(n-2)\widehat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(s_k - \overline{s})^2}{\sum_{i=1}^{n}(s_i - \overline{s})^2}}$$

Solving yields:

$$s_k = \overline{s} + \sqrt{\left[\frac{\left(L - \widehat{Y}_{s_k}\right)^2}{t^2\widehat{\sigma}^2} - 1 - \frac{1}{n}\right]\sum_{i=1}^{n}(s_i - \overline{s})^2}$$

where t is the $100(1 - \alpha)$ percentile of a t distribution with $n - 2$ degrees of freedom.

Problem 2 is simpler. Presume that there is a desired warranty time, call it s_w, and we want to know how likely it is that the product will have an adequate value of Y at that time. If the predicted value of Y at s_w is $\widehat{Y}_{s_w}$, then the lower confidence limit for the predicted value at time s_w is:

$$\widehat{Y}_L = \widehat{Y}_{s_w} - t^*\widehat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(s_w - \overline{s})^2}{\sum_{i=1}^{n}(s_i - \overline{s})^2}}$$

Setting $\widehat{Y}_L = L$ and solving for t^* gives:

$$t^* = \frac{\widehat{Y}_{s_w} - \widehat{Y}_L}{\widehat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(s_w - \overline{s})^2}{\sum_{i=1}^{n}(s_i - \overline{s})^2}}}$$

The probability that Y will not fall below Y_L before time s_w, $Pr\{Y \geq Y_L | s_w\}$, is $Pr\{T \leq t^* | n - 2\}$, the probability that T, a Student's t random variable with $n - 2$ degrees of freedom, will be less than t^*.

Generally, design problems cannot afford to wait for product to be put on test until the expiration or warranty time has elapsed. Thus, some form of accelerated test is generally required. Presume that the accelerating parameter is temperature. The question is how to determine the temperature and the degree to which acceleration is achieved at that temperature. The methods for determining the Arrhenius parameters described earlier could be employed. However, since the response variable, Y, is continuously valued, and is assumed here to be monotonic and linear with respect to time, there is another possible approach. Perform the test at a fixed set of times, say $s_1, s_2, \ldots s_n$, at two temperatures, P_1 and P_2, both of which are greater than the "nominal" temperature, P_0 (typically $P_0 = 25\ ^\circ\text{C}$), and $P_2 > P_1$. Obtain slope estimates for each set of data. Call the estimates b_1 (at temperature P_1) and b_2 (at temperature P_2). An estimate of the acceleration rate for a temperature difference of $\Delta_P = P_2 - P_1$ is $k_{\Delta P} = b_2 / b_1$. If the EAS desires to simulate T_0 time units with $T_a < T_0$ time units, she or he would require $k_a \approx T_0 / T_a$. Let

$\Delta_a = P_a - P_0$ be the temperature difference from nominal required to achieve acceleration k_a. Then:

$$\frac{\Delta_a}{\Delta_P} = \frac{k_a}{k_{\Delta P}} \Rightarrow \Delta_a = \frac{k_a}{k_{\Delta P}} \Delta_P$$

So, to simulate T_0 time units at nominal temperature P_0, test over T_a time units at temperature $P_a = P_0 + \Delta_a$. The number of time points at which to measure Y should be at least 3, 2 more than the highest order term in the model, but 4 or 5 would be better. For the purposes of assessing the adequacy of the model (which we have assumed was first-order), replication at each time point in the form of multiple units measured, is highly recommended.

For the EAS, the values of s_k or t^* depend on the product, which is in turn dependent upon the chemical or other components. If prototype products are formulated in a designed experimental fashion, then either s_k or t^* could act as response variables. In that way, a desired prototype formulation could be found.

In the shelf life discussion so far, we have only dealt with a case where the quality response variable is monotonically decreasing, and that there was a lower limit of acceptability. The case where Y is monotonically increasing, with an upper limit of acceptability, is analogous.

The purpose of this section was to provide some ideas concerning shelf-life testing, and to encourage the use of designed experiments in order to achieve a desired shelf-life with some stated level of probability. Shelf-life is a fairly broad topic, and those who are interested are encouraged to read the book by Professor Chow (2007)

Key Points

- Reliability is the probability that something will not fail before a given time.
- Time-to-event variables give rise to reliability.
- Reliability models stem from the same first-order differential equation describing radioactive decay and chemical reaction kinetics.
- The cdf, reliability, and hazard functions for a time-to-event variable are all related; one can be derived from the others.
- When the hazard rate is a constant, time-to-event has an exponential distribution.
- Accelerated life tests involve a model for assessing the degree to which the hazard rate is affected by "accelerated" conditions (most notably temperature).
- Shelf life or stability is often assessed using regression models.

Exercises and Questions

1. The Weibull cumulative distribution function can be expressed as: $F(t) = 1 - exp\left(-(\alpha t)^{\beta}\right)$ where α and β are called the scale and shape parameters, respectively. Derive the hazard rate function, $h(t)$.
2. Could you use the cumulative hazard rate function, $H(t)$, of the Weibull distribution together with sample life test data to estimate the values of α and β via least squares regression? How would you do it?

3. Given a set of uncensored life test data, would you rather approximate $H(t)$ with a low-order polynomial or assume that the time-to-failure has a Weibull distribution?
4. A failure process has a failure rate that follows the Arrhenius law, with $B = 6700$. At $P_0 = 25 + 273 = 298°\text{K}$, what temperature would you recommend to accelerate the failure rate 10 times?
5. At $P_0 = 298°\text{K}$, what is the failure rate if $A = 1.0\text{E}05$ (per hour)?

Chapter 12
Some Bayesian Concepts

Bayesian statistical methods are based on Bayes' theorem, which was described in Chap. 1. Suppose X represents a continuously-valued random variable, and $f(x|\boldsymbol{\theta})$ is its density function given parameter $\boldsymbol{\theta}$. In the Bayesian world, the parameter θ is also treated as a random variable, with a density function $g(\theta)$. This density is referred to as the prior density for θ, inasmuch as it is formulated prior to obtaining any observations of X. The idea is that $g(\theta)$ represents our prior belief about the likelihood that θ takes on a value in any particular range. Generally, $g(\theta)$ is also a function of some other parameters, which we will call hyperparameters, whose values are chosen to reflect the prior belief about the possible range of values for θ. The observation of X, call it x, is assumed to be dependent on the value of θ. The dependency is expressed as a likelihood function, symbolized by $L(x|\theta)$. Once the data, x, are observed, the Bayesian would like to update his or her belief concerning the probability that the unknown parameter, θ, falls in any particular range. The updated belief is expressed as a conditional density function, called the posterior density, and is expressed as $g(\theta|x)$. Bayes' theorem provides a method for deriving the posterior density given the prior density and the likelihood function:

$$g(\theta|x) = \frac{L(x|\theta)g(\theta)}{\int_{-\infty}^{+\infty} L(x|\tau)g(\tau)d\tau}$$

Conjugacy is a condition that greatly simplifies computations. A likelihood function and a prior distribution are said to be a conjugate pair if the resulting posterior distribution is of the same form as the prior. For conjugate pairs, the posterior distribution has hyperparameters whose values are generally a closed-form function of the prior hyperparameters and the data. Generally this relationship is the reason why conjugacy greatly simplifies computations.

We will focus on two particular but hopefully useful conjugate cases. The first is the case where data are binomially distributed with a beta distribution as the prior for the success probability parameter. The second case has normally distributed data, where the mean parameter has a normal prior. A very complete exposition of Bayesian methods is given in Gelman et al. (1997).

S.A. Pardo, *Empirical Modeling and Data Analysis for Engineers and Applied Scientists*,
DOI 10.1007/978-3-319-32768-6_12

Binomial Data with Beta Prior

Let X be a binomially distributed random variable with probability mass function:

$$L(x|p,n) = \binom{n}{x} p^x (1-p)^{n-x}$$

Furthermore, suppose the parameter p is unknown, and n represents the sample size to be drawn. Then we might choose as a prior distribution for p a beta with hyperparameters α and β, i.e.,

$$g(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

Then, given the data x = number of "successes" observed in a sample of size n, the posterior probability density function of p is given by:

$$g(p|x,n) = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)} p^{\alpha+x-1}(1-p)^{\beta+n-x-1}$$

In other words, the posterior distribution for p is also beta, with hyperparameters

$$\alpha + x \text{ and } \beta + n - x$$

The prior hyperparameters, α and β, can be thought of as the best guess for percent success (α) and percent failure (β). That is, if the EAS chooses α and β so that their sum adds to 100, then she or he could choose their respective values to represent the best guess, prior to getting any data, on the probability of success and failure. It is not necessary to have $\alpha + \beta = 100$; it is just a potentially convenient way to quantify prior belief.

The expected value of a beta-distributed random variable (i.e., p) is given by:

$$E[p|\alpha,\beta] = \frac{\alpha}{\alpha+\beta}$$

That means the expected value for the posterior distribution is:

$$E[p|\alpha+x,\beta+n-x] = \frac{\alpha+x}{\alpha+\beta+n}$$

Note that unlike the normal distribution, the expected value for the beta is not the value of p that maximizes the density function.

Suppose an EAS has some belief that a particular part design has a 90 % chance of performing properly. She chooses a beta prior for p, the probability of proper performance, and sets $\alpha = 90$, and $\beta = 10$. Figure 12.1 shows the prior density for p.

The value of p that maximizes the density is 0.908, which is close, but not identical, to the expected value of 0.90 (0 %).

She performs a test with $n = 100$ parts, and finds that only $x = 75$ perform adequately. The posterior expectation is:

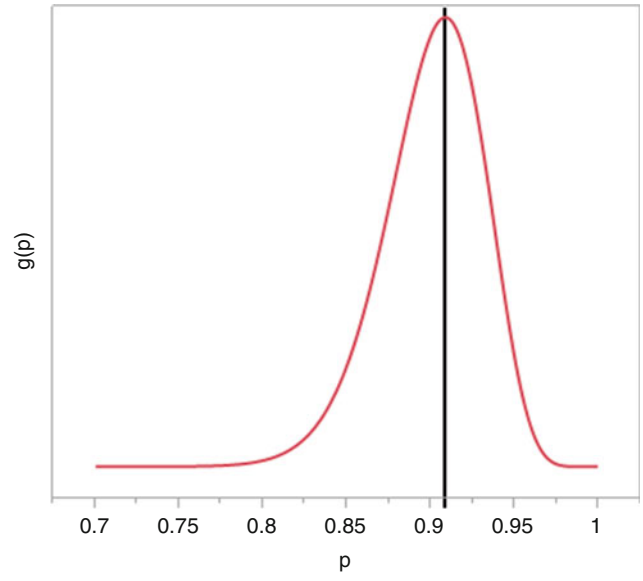


Fig. 12.1 Prior beta density for p

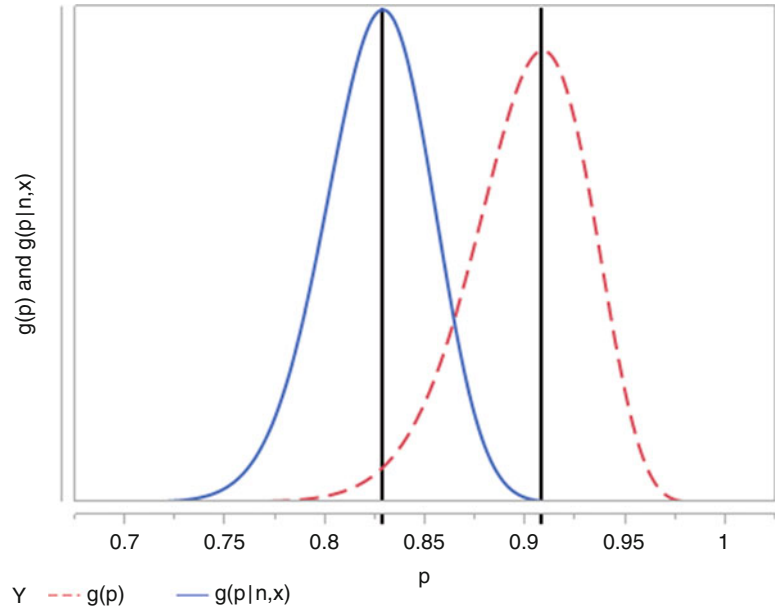


Fig. 12.2 Prior and posterior densities for p

$$E\left[p\middle|\alpha+x,\beta+n-x\right]=\frac{\alpha+x}{\alpha+\beta+n}=\frac{90+75}{90+10+100}\approx 0.825$$

The prior and posterior densities are shown in Fig. 12.2.

The value of p that maximizes the posterior density is 0.828, and not the expected value of 0.825.

The prior probability that p is at least 0.90 is approximately 53.55 %. The posterior probability is only about 0.07 %. It appears that this design is not adequate, assuming that the EAS was hoping there was at least a 90 % chance that the part would perform properly.

Normal Data with Normal Prior

The normal likelihood for X given μ and σ is:

$$L[x|\mu,\sigma] = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-0.5\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

That is, the likelihood for an individual observation is the normal density function. For a random sample of n values, the likelihood is:

$$L[x_1,\ldots,x_n|\mu,\sigma] = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2\sigma}\sum_{i=1}^{n}x_i^2 + \frac{n\mu\bar{x}}{\sigma} - \frac{n\mu^2}{2\sigma}\right)$$

Suppose that σ is known (at least, for now). If the prior density chosen for the unknown parameter, μ, is also a normal density of the form:

$$g(\mu|\mu_0,\tau_0) = \frac{1}{\tau_0\sqrt{2\pi}}\exp\left(-0.5\left(\frac{\mu-\mu_0}{\tau_0}\right)^2\right)$$

Then the posterior parameters after observing n values of X would be:

$$\mu_n = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{x}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}$$

$$\tau_n^2 = \frac{1}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}$$

$\bar{x}$ is the arithmetic average of the n values of X.

The posterior density of μ would then be:

$$g(\mu|\mu_n,\tau_n) = \frac{1}{\tau_n\sqrt{2\pi}}\exp\left(-0.5\left(\frac{\mu-\mu_n}{\tau_n}\right)^2\right)$$

As an example, consider an electrical motor. It is guessed that this motor can generate 44.5 Newton-meters (Nm) of torque. The EAS recognizes that each unit may vary the amount of torque even if the same voltage is applied. He has found a record of torque measurements where the reported standard deviation, σ, was 1.105 Nm. The mean torque values were also reported, with a mean value of $\mu_0 = 44.50$ Nm. The standard deviation of these means was $\tau_0 = 0.300$ Nm. He plans to make torque measurements on some new units, but he does not want to disregard historical data. He decides that the mean torque should have a prior density assigned, with $\mu_0 = 44.50$ Nm and $\tau_0 = 0.300$ Nm. His experiment with n = 10 units yields $\bar{x} = 47.56$ Nm. Using these data, he updates his prior hyperparameters:

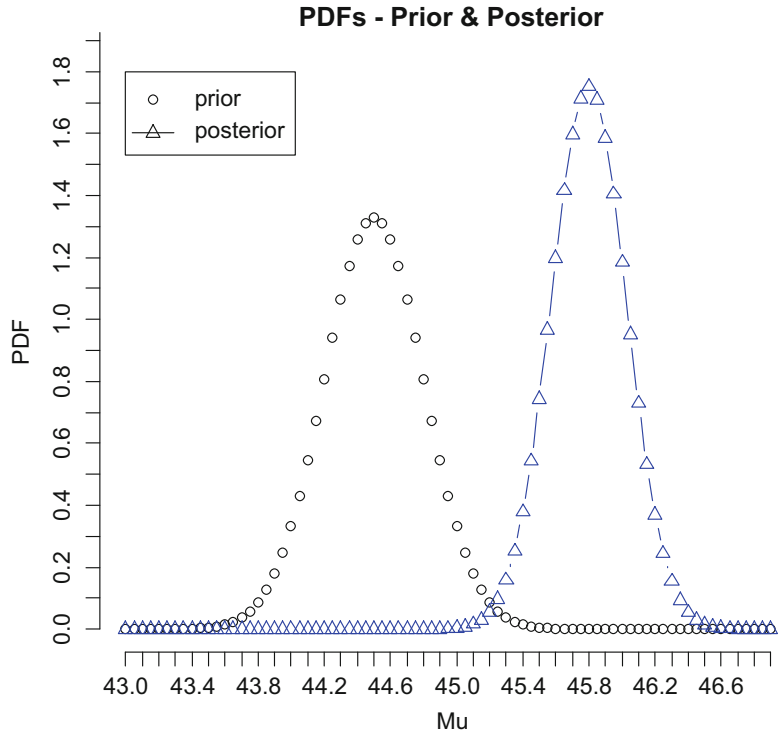


Fig. 12.3 Prior and posterior probability density functions for μ

$$\mu_{n=}\frac{\frac{1}{\tau_0^2}\mu_0+\frac{n}{\sigma^2}\overline{x}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\approx 45.80$$

$$\tau_n^2=\frac{1}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\approx 0.228^2\approx 0.05198$$

Prior to obtaining any data, the EAS was about 95 % certain that the value of μ was somewhere between $\mu_0 - 1.96\tau_0 = 44.50 - 1.96(0.300) \approx 43.91$ Nm and $\mu_0 - 1.96\tau_0 = 44.50 + 1.96(0.300) \approx 45.09$ Nm. After obtaining the data, he now believes with 95 % certainty that μ is somewhere between $\mu_n - 1.96\tau_n = 45.80 - 1.96(0.228) \approx 45.35$ Nm and $\mu_n - 1.96\tau_n = 45.80 + 1.96(0.228) \approx 46.25$ Nm. These intervals are called "credible intervals" (Gelman et al. 1997), and are a sort of analog to the classical frequentist confidence interval concept. Figure 12.3 illustrates the prior and posterior densities of μ. Figure 12.4 gives the R code that was used to generate the graph.

When σ is Unknown

The previous analyses presumed that somehow the parameter σ was known. While it may be possible that historical data might give rise to a more or less trustworthy value for σ, it often will not be the case. It turns out that determining a posterior distribution for μ when σ is also unknown is more complex. Gelman, et al., give a sort of way to get a reasonable approximation. First, assume a prior distribution for σ. In particular, suppose that $\frac{1}{\sigma^2}$ has a gamma distribution with hyperparameters ν_0 and $\beta_0 = \frac{\nu_0\sigma_0^2}{2}$. The posterior hyperparameters, given the data:

```
setwd("H:\\Personal Data\\Experimentation for Design & Validation\\Data &
Analyses\\")
df1 <- read.csv("20141127 example 12.2 priors and data.csv")
# Inputs:
# Mu0 - prior mean hyperparameter
# Tau0 - prior sd hyperparameter
# sigma - sd of measurements
# n - sample size
# xbar - observed arithmetic average
#
attach(df1)
mu <- c()
pdf0 <- c()
pdf1 <- c()

denom <- (1/Tau0**2) + (n/sigma**2)
#
# compute posterior hyperparameters:
#
Mun <- (Mu0*(1/Tau0**2) + xbar*(n/sigma**2))/denom
Taun <- sqrt(1 / denom)
mulow <- min(Mu0-5*Tau0,Mun-5*Tau1)
muhigh <- max(Mu0+5*Tau0,Mun+5*Tau1)
mu <- seq(mulow, muhigh,0.05)
pdf0 <- dnorm(mu,Mu0,Tau0)
pdfn <- dnorm(mu,Mun,Taun)
height <- max(max(pdf0),max(pdfn))
incr <- 0.1
#
# col = 1  Black
# col = 2  Red
# col = 3  Green
# col = 4  Blue
# col = 5  Light Green
# col = 6  Violet
# col = 7  Yellow
# col = 8  Light Grey
#

plot(mu,pdf0,axes=F,type="n",main="PDFs - Prior &
Posterior",xlim=c(mulow,muhigh),ylim=c(0,height+incr),xlab="Mu",ylab="PDF")
axis(side=2,at=seq(0,height+6,0.1)) #defines y-axis
axis(side=1,at=seq(mulow,muhigh,0.1)) #defines x-axis
points(mu,pdf0,pch=1,type="b",lty=2,col=1)
points(mu,pdfn,pch=2,type="b",lty=1,col=4)
legend(43,1.8,legend=c("prior","posterior"),pch=c(1,2),lty=c(2,1))
```

Fig. 12.4 R code for generating the prior and posterior density graph

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

are:

$$\nu_n = \nu_0 + n$$

$$\beta_n = \frac{2\beta_0 + n\widehat{\sigma}^2}{\nu_0 + n}$$

You may have noticed a small fly in this ointment, namely that the posterior parameters depend on knowing μ. It turns out that the joint distribution of μ and σ can be derived in a closed form. However, another method may be used to estimate the posterior expected value of μ. The method, referred to as the Monte Carlo Markov Chain (MCMC) method, is described in Gelman et al. (1997). The idea is to randomly draw a value from the prior distribution of $\frac{1}{\sigma^2}$. Then use this value to randomly generate a value from the posterior distribution of μ. Then, in turn, use this randomly generated value of μ to randomly generate a value of $\frac{1}{\sigma^2}$ from its posterior distribution. Continually iterate, updating the hyperparameters at each iteration, for N times.

Our EAS decides to run an MCMC program, with N = 10,000, as illustrated in Fig. 12.5. The output values of the prior and posterior hyperparameters, together with the sample mean and standard deviation of the raw data, are given in Table 12.1. The prior and posterior density functions are plotted in Fig. 12.6.

In this relatively simple case, the data have overwhelmed the prior guess; the posterior hyperparameter μ_n is equal to the sample mean, rounding to two decimal places of numerical precision.

How can Bayesian methods be used to aid in design? One idea is to have a set of prior hyperparameters for each run in a designed experiment. After obtaining data, update the hyperparameters, and compute credible intervals corresponding to each run for the parameter of interest. As a simple example, consider a single continuously valued response variable with a single continuously valued regressor.

The conditional expectation for the response, y, is:

$$E[y|x] = \beta_0 + \beta_1 x$$

The response variable is usually assumed to have a normal distribution with this conditional expectation and variance σ^2. Thus, the likelihood function is based on the normal density:

$$y(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{y - \beta_0 - \beta_1 x}{\sigma}\right)^2\right)$$

The regression coefficients, β_0 and β_1, are assumed to have a bivariate normal prior distribution with mean vector:

$$\boldsymbol{\gamma}_0 = \begin{bmatrix}\gamma_0 \\ \gamma_1\end{bmatrix}$$

and covariance matrix:

$$\sigma^2[\boldsymbol{X}'\boldsymbol{X}]^{-1}$$

```
setwd("H:\\Personal Data\\Experimentation for Design & Validation\\Data &
Analyses\\")
df1 <- read.csv("20141128 example 11.3 data.csv")
# Inputs:
# Mu0 - prior mean hyperparameter
# Tau0 - prior sd hyperparameter
# sigma - sd of measurements
# x = vector of data values
#
attach(df1)
mu <- c()
pdf0 <- c()
pdf1 <- c()
N <- 10000
Mu0 <- 44.5
Tau0 <- 0.300
nu0 <- 100
sig0 <- 1
beta0 <- (nu0*sig0**2)/2
sigma <- 1 / rgamma(1,shape=nu0,scale=beta0)
xbar <- mean(x)
n <- length(x)
nun <- nu0 + n
for (i in 1:N) {
   denom <- (1/Tau0**2) + (n/sigma**2)
#
# compute posterior hyperparameters:
#
  Mun <- (Mu0*(1/Tau0**2) + xbar*(n/sigma**2))/denom
  Taun <- sqrt(1 / denom)
  murand <- rnorm(1,mean=Mun,sd=Taun)
  ssx <- (x - murand)**2
  sig2ml <- mean(ssx)
  nu0 <- nu0 + n
  beta0 <- (2*beta0 + n*sig2ml) / nu0
  sigma <- 1 / rgamma(1,shape=nu0,scale=beta0)

  }
```

Fig. 12.5 MCMC program (R code)

```
mulow <- min(Mu0-5*Tau0,Mun-5*Tau0)
muhigh <- max(Mu0+5*Tau0,Mun+5*Tau0)
mu <- seq(mulow, muhigh,0.05)
pdf0 <- dnorm(mu,Mu0,Tau0)
pdfn <- dnorm(mu,Mun,Taun)
height <- max(max(pdf0),max(pdfn))
incr <- 0.1
#
# col = 1  Black
# col = 2  Red
# col = 3  Green
# col = 4  Blue
# col = 5  Light Green
# col = 6  Violet
# col = 7  Yellow
# col = 8  Light Grey
#

plot(mu,pdf0,axes=F,type="n",main="PDFs - Prior &
Posterior",xlim=c(mulow,muhigh),ylim=c(0,height+incr),xlab="Mu",ylab="PDF")
axis(side=2,at=seq(0,height+6,0.1)) #defines y-axis
axis(side=1,at=seq(mulow,muhigh,0.1)) #defines x-axis

points(mu,pdf0,pch=1,type="b",lty=2,col=1,cex=0.3)
points(mu,pdfn,pch=2,type="b",lty=1,col=4,cex=0.3)
legend(43,10,legend=c("prior","posterior"),pch=c(1,2),lty=c(2,1))
```

Fig. 12.5 (continued)

Table 12.1 MCMC output

MCMC parameters	
nu0	100
beta0	50
sig0	1
Mu0	44.5
Tau0	0.300
Mun	47.56
Taun	0.0037
xbar	47.56
sd(x)	0.167

where:

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

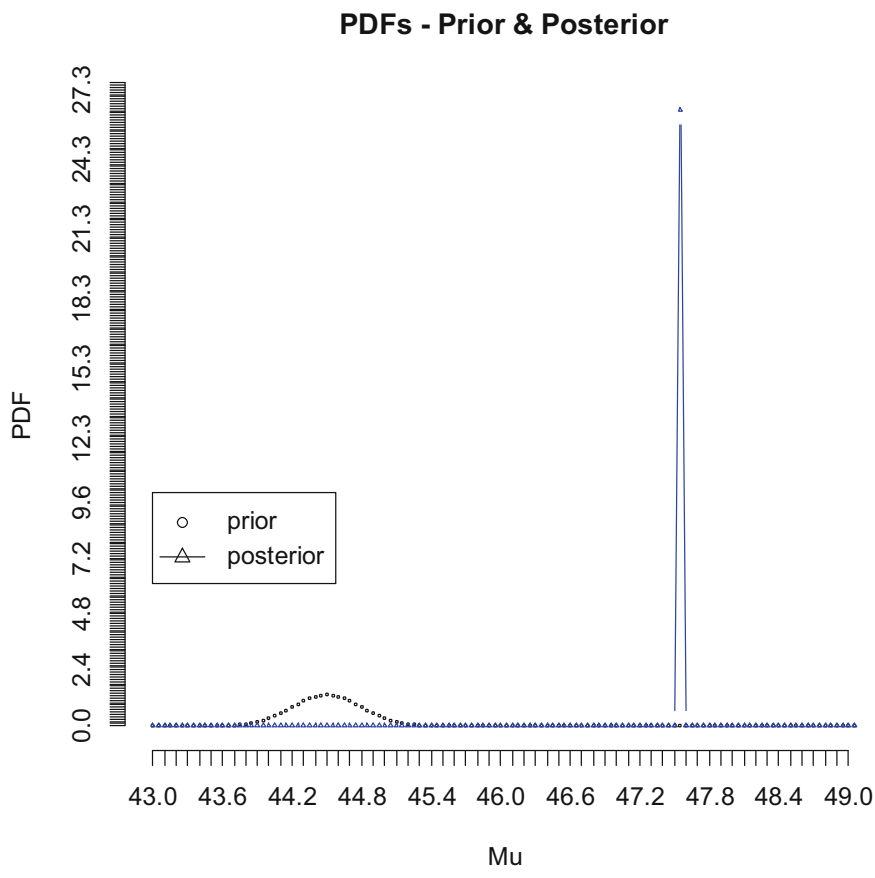


Fig. 12.6 Prior and posterior densities from MCMC program

Here it is assumed that σ is known. The conditional updating formulae (Box and Tiao 1973) for the posterior mean vector of the coefficients is given by:

$$\boldsymbol{\gamma_n} = \frac{1}{2}[\boldsymbol{X}'\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{X}'\boldsymbol{X}\boldsymbol{\gamma}_0$$

with $\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ being the vector of n observed values of the response.

An MCMC procedure could be used to update the posterior parameters for the coefficients, using the same prior and posterior conditional gamma conjugate pair for $\frac{1}{\sigma^2}$, with μ replaced with:

$$\mu(x_i) = \gamma_0 + \gamma_i x_i$$

so that:

$$\widehat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \mu(x_i))^2$$

Typically, the estimate $\widehat{\sigma}^2$ from a regression model would use $n - k$ as the denominator, where $k =$ number of model coefficients (in this case, $k = 2$).

If data are obtained sequentially, the EAS can use a recursive updating formula for the regression posterior hyperparameters, given by Judge et al. (1985). The updating formulae, given a new pair of values, x_{n+1}, y_{n+1}, is:

$$\boldsymbol{\gamma}_{n+1} = \boldsymbol{\gamma}_n + \frac{\boldsymbol{S}_n \boldsymbol{x}_{n+1}\left(y_{n+1} - \boldsymbol{x}'_{n+1}\boldsymbol{\gamma}_n\right)}{1 + \boldsymbol{x}'_{n+1}\boldsymbol{S}_n\boldsymbol{x}_n}$$

and

$$\boldsymbol{S}_{n+1} = \boldsymbol{S}_n - \frac{\boldsymbol{S}_n \boldsymbol{x}_{n+1}\boldsymbol{x}'_{n+1}\boldsymbol{S}_n}{1 + \boldsymbol{x}'_{n+1}\boldsymbol{S}_n\boldsymbol{x}_n}$$

where

$$\boldsymbol{x}'_{n+1} = [1\ x_{n+1}] \text{ and}$$

$$\boldsymbol{S}_n = \frac{1}{(n+1)\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n+1 \end{bmatrix}$$

As an alternative, somewhat unorthodox, model, suppose that the prior hyperparameters are the slope, β_1, and intercept, β_0, of a linear model, with an unknown noise variance:

$$\mu \sim N\left(\beta_0 + \beta_1 x,\ \tau_0^2\right)$$

and:

$$\frac{1}{\sigma^2} \sim Gamma\left(\nu_0, \gamma_0 = \frac{\nu_0 \sigma_0^2}{2}\right)$$

With a normal likelihood function, the conditional posterior distribution of μ, given σ, is also normal, with posterior hyperparameters:

$$\mu_n(x) = \frac{\frac{1}{\tau_0^2}(\beta_0 + \beta_1 x) + \frac{n}{\sigma^2}\overline{y}(x)}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}$$

$$\tau_n^2 = \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}$$

The values of β_0 and β_1 could be obtained by least squares either using previously gathered data or even a partial subset of the data just gathered (i.e., partition the data into a "training set" and a "test set"; the training set could be used to obtain prior values for the hyperparameters). Of course, the parameters β_0 and β_1 could also have prior distributions assigned with their own hyperparameters, making the model complex, and a good candidate for MCMC methods.

Key Points

- Bayes' Theorem provides a means of incorporating prior information about parameters together with empirical observations.
- Bayesian methodology quantifies information about parameters using probability distributions that describe the degree of uncertainty about the parameters.
- The parameter distribution describing the uncertainty about the parameter before any new data are gathered is called the prior distribution.
- The probability function that describes the chance of observing particular values given particular parameter values is called the likelihood function .
- The posterior distribution for parameters describes the uncertainty about those parameters after data have been gathered; it is computed using the prior distribution and the likelihood function.
- If the posterior distribution can be derived analytically, and has the same parametric form as the prior, then the prior and associated likelihood function are said to be a conjugate pair.
- Even when a likelihood function and prior are not conjugate, they can be used to compute the posterior distribution. A technique for using computer simulation with non-conjugate pairs is called Monte Carlo Markov Chains (MCMCs).

Exercises and Questions

1. Suppose a binomial likelihood function is used together with a uniform prior distribution, namely:

$$g(p) = \frac{1}{b-a}, \ \forall p$$

   where a and b are constants, $b > a$. What would the posterior distribution for p look like? Are these a conjugate pair?
2. Discuss the appropriateness of Bayesian methods for utilizing experimental data to aid in design. Consider the formulation of prior distributions, and the possible uses of MCMC methods.

Chapter 13
Validation and Verification

The terms validation and verification often are used in very specific ways (Nash and Wachter 2007). Here we will use the terms somewhat loosely, more or less adhering to the following definitions:

Validation: demonstrating that within some tolerance ranges for product or process features, the values of response variables (measures of goodness) are acceptable with some associated probability measure.

Verification: demonstrating that a predictive model predicts new response values within acceptable range of error.

Verification

Beginning with the epoxy example in Chap. 4, we have admonished the EAS to verify predictions by obtaining some new experimental results. We did not suggest any formal verification process or procedure. If the predicted response is $\widehat{y}(\boldsymbol{x})$, where $\boldsymbol{x}$ is a vector of input variable values, and $\overline{y}_v(\boldsymbol{x})$ is the average of m new response values obtained to verify the model, then the EAS would want the difference $\overline{y}_v(\boldsymbol{x}) - \widehat{y}(\boldsymbol{x})$ to be within some desired limits. Using the data from the model fitting, we would have an estimate of the variability in this difference, namely:

$$Var(\overline{y}_v(\boldsymbol{x}) - \widehat{y}(\boldsymbol{x})) = \frac{\widehat{\sigma}^2}{m} + se^2(\widehat{y}(\boldsymbol{x}))$$

An approximate 95 % confidence interval for the difference would be:

$$\overline{y}_v(\boldsymbol{x}) - \widehat{y}(\boldsymbol{x}) \pm t_{.975}(n+m-(k+1))\sqrt{\frac{\widehat{\sigma}^2}{m} + se^2(\widehat{y}(\boldsymbol{x}))}$$

The symbol *se* refers to the estimated standard error of the predicted value, and n is the sample size used to fit the model. The symbol $t_{.975}(n + m - (k + 1))$ represents the 97.5th percentile of s Student's t distribution with $n + m - (k + 1)$ degrees of freedom. If this confidence interval falls within a desirable range, then the model is verified with 95 % confidence. This is a very simplistic verification process, but it may be sufficient. More sophisticated methods might include saving some data as a "test" set, and not including them in the fitting process. Other methods might include an iterative

S.A. Pardo, *Empirical Modeling and Data Analysis for Engineers and Applied Scientists*,
DOI 10.1007/978-3-319-32768-6_13

process, where random subsets of data are selected to fit, and then the differences between predicted and actual responses could be computed. In each iteration, the mean square error (*MSE*) could be computed for the "test" set of m values ($m < n$):

$$MSE = \frac{1}{m}\sum_{j=1}^{m}\left(y_i - \hat{y}_j\right)^2$$

The *MSE* values from all the iterations could be averaged, and the square root of the average *MSE* could be computed (call is root mean square, or *RMSE*). If this number is a "small enough" percentage of the average actual response values, then the model might be considered as verified.

Validation

We will use the word "validation" to mean validating that the product or process performs adequately, once it has been designed and manufactured. Of primary importance is defining the range of response values that would be considered "adequate". The next question is by how much can any of the product features vary and still provide with sufficient probability response values that will fall within the adequate range. One possible means is to interpolate data from an experiment used to optimize the product design, by finding the ranges of factor or regressor values that will result in predicted values falling within the adequate performance range. As an example, consider the hip replacement example in Chap. 10. The desired reliability at 20 years of life was 95 %. The 95 % confidence interval for reliability at 20 years for design P5 was (95.85 %, 98.26 %). Since, in this case, the features, or factors, were discrete, the EAS may consider this design "validated", at least with respect to reliability. For continuously valued design features with a continuous response, the methods described in Chap. 8, *Range Finding*, would be appropriate for product validation. Note that a confirmatory experiment, centered around the "optimal" design point, should be executed. By way of simple example, suppose there is only a single continuously valued feature (regressor). Suppose further that a first-order model is adequate. Suppose x^* represents the desired design point. Furthermore, suppose that the values of x that describe the acceptable operating range are x_{LL} and x_{UU}, as defined in Chap. 8. The EAS performs a confirmatory experiment with $x = x_{LL}$ and $x = x_{UU}$, and observes n values of the response at each of these points. She then computes the means and standard deviations y for each group. The desired value for the response at optimal point x^* is y^*, and the limits of acceptability are $y^* \pm \delta$. Compute the confidence range of a future value for the response at each of x_{LL} and x_{UU}, namely:

$$L = \overline{y}_{LL} + ts\sqrt{1+\frac{1}{n}} \quad \text{and} \quad U = \overline{y}_{UU} - ts\sqrt{1+\frac{1}{n}}$$

If $L \geq y^* - \delta$ and $U \leq y^* + \delta$, then the product is valid with respect to y, in terms of x. The value of t is the $100(1 - \alpha)$ percentile of s Student's t distribution with $n - 1$ degrees of freedom. The ideas are first, that if the value of the response was truly $y^* - \delta$ or $y^* + \delta$, the EAS would be satisfied with performance, and that the true performance could not be both $y^* - \delta$ and $y^* + \delta$ simultaneously.

The same notions apply when there is an optimal vector of feature/setting values, $\boldsymbol{x}^*$, but the computation of limit vectors, $\boldsymbol{x}_{LL}$ and $\boldsymbol{x}_{UU}$, is more complicated.

Sometimes validation is based on a discrete random variable, b, with a Bernoulli distribution, i.e., the variable b equals 1 if an observation is a "success" and 0 otherwise. There may be a minimum

desired probability of "success", say p^*. If n independent Bernoulli observations, b_i, i = 1, n, are obtained, and

$$X = \sum_{i=1}^{n} b_i$$

then X, the number of successes out of n independent Bernoulli trials, has a binomial distribution. Thus, a critical value, or acceptance criterion, x_c, can be found for X. That is, x_c is chosen so that:

$$Pr\{X \geq x_c | p^*, n\} = \sum_{k=x_c}^{n} \binom{n}{k} (p^*)^k (1-p^*)^{n-k} = 1 - \sum_{k=0}^{x_c-1} \binom{n}{k} (p^*)^k (1-p^*)^{n-k} \approx 1 - \alpha$$

To be conservative, one should choose the critical value, x_c, such that:

$$1 - \alpha = sup\ Pr\{X \geq x_c | p^*,\ n\}$$

The term "*sup*" is an abbreviation for "supernum", or least upper bound. The reason for specifying that $1 - \alpha$ should be the least upper bound on the probability of "passing" the test when the actual probability of "success" is p^* is due to the discrete nature of the variable X. For example, if $n = 100$, $p^* = 0.90$, and we want $1 - \alpha = 0.95$. If we choose $x_c = 85$, then

$$Pr\{X \geq 85 | p^* = 0.90, n = 100\} \approx 0.9601$$

If we choose $x_c = 86$, then

$$Pr\{X \geq 86 | p^* = 0.90, n = 100\} \approx 0.9274$$

Since we want $1 - \alpha = 0.95$ to be the least upper bound on the probability of "passing", then we should choose $x_c = 86$ instead of 85. That is, with $n = 100$ and $p^* = 0.90$, it is not possible to find an integer critical value such that the probability of passing is exactly 0.95.

Such a criterion can be thought of as the critical value for testing the Noninferiority hypothesis (Pardo 2014):

$$H_0 : Pr\{\text{success}\} < p^* \text{ versus the alternative } H_a : Pr\{"success"\} \geq p^*.$$

In other words, if $X \geq x_c$, then reject the null hypothesis, H_0, in favor of H_a.

Validation is usually a multivariate problem, inasmuch as any product will have more than one key response variable measuring the product's performance. That is, there could be k response variables, $Y^{(1)}, Y^{(2)}, \ldots, Y^{(k)}$, that describe the performance of a product or system. Furthermore, suppose all the $Y^{(j)} > 0$, and that they are dependent on the same vector or regressors, $\boldsymbol{x}$, which represent the product or system features. One strategy is to map the multiple responses into "desirability" functions, which can then be mapped into an overall desirability. There are three forms of desirability mappings: lowest-is-best, target-is-best, and highest-is-best. For each response variable, $Y^{(j)}$, let $Y_{low}{}^{(j)}$ be a value a little less than the minimum value of $Y^{(j)}$, and $Y_{high}{}^{(j)}$ be a value a little greater than the maximum. Also, suppose that if there is a target value, it is represented as $Y_{targ}^{(j)}$, $Y_{low}^{(j)} < Y_{targ}^{(j)} < Y_{high}^{(j)}$. Then the lowest-is-best desirability is given by:

$$d^{(j)} = \left(\frac{Y_{high}^{(j)} - Y^{(j)}}{Y_{high}^{(j)} - Y_{low}^{(j)}} \right)^r$$

The highest-is-best desirability is:

$$d^{(j)} = \left(\frac{Y^{(j)} - Y_{low}^{(j)}}{Y_{high}^{(j)} - Y_{low}^{(j)}} \right)^r$$

The target-is-best desirability is:

$$d^{(j)} = \left(1 - \frac{\left| Y^{(j)} - Y_{targ}^{(j)} \right|}{max\left(Y_{high}^{(j)} - Y_{targ}^{(j)}, Y_{targ}^{(j)} - Y_{low}^{(j)} \right)} \right)^r$$

The overall desirability, D, is the geometric mean of the $d^{(j)}$:

$$D = \left(d^{(1)} d^{(2)} \cdots d^{(k)} \right)^{1/k}$$

The parameter $r > 0$ is selected by the EAS; if $r = 1$, then desirability is linear; if $r < 1$, the desirability is convex (less discriminating); $r > 1$ then it is concave (more discriminating). The maximum possible value for D is 1. The EAS may optimize simultaneously with respect to all the response variables by finding the values of the regressors that maximize D. The idea is to suppose that $\boldsymbol{x}^*$ represents the vector of regressors that maximize D (subject of course to whatever constraints there may be on $\boldsymbol{x}$), and D^* is the optimal value of overall desirability. Also let $\boldsymbol{\delta}$ be a vector of tolerances for the components of $\boldsymbol{x}^*$. Then suppose the two vectors $\boldsymbol{x}^* \pm \boldsymbol{\delta}$ are used to find values of D based on an experimental run with the regressor values $\boldsymbol{x}^* - \boldsymbol{\delta}$ and $\boldsymbol{x}^* + \boldsymbol{\delta}$. Furthermore, suppose the EAS defines a range of tolerance for desirability, namely that predicted desirability off the optimal point must be no less than $D^* - \rho$, where ρ is some predetermined quantity. Let the two values $\widehat{D}_{-\delta}$ and $\widehat{D}_{+\delta}$ represent the predicted (average) overall desirabilities at each of the two "suboptimal" points. Then the conditions for validation are:

$$\widehat{D}_{-\delta} + t\frac{s_{D_{-\delta}}}{\sqrt{n}} \geq D^* - \rho \quad \text{and} \quad \widehat{D}_{+\delta} + t\frac{s_{D_{+\delta}}}{\sqrt{n}} \geq D^* - \rho$$

where $s_{D_{-\delta}}$ and $s_{D_{+\delta}}$ are the standard deviations of overall desirabilities $D_{-\delta}$ and $D_{+\delta}$, n is the sample size (not required to be the same for both $-\delta$ and $+\delta$ cases), and t is the $100(1 - \alpha)$ percentile of a Student's t distribution with $n - 1$ degrees of freedom.

Validation is a complex process that involves much more than the analyses of experimental data. The techniques mentioned in this chapter, and in Chap. 8, will potentially aid in minimizing the amount of experimentation necessary for validation. More details on the ideas of the tests for validation can be found in Pardo (2014).

Key Points

- Validation involves demonstrating that a product or system performs adequately as long as input features are within tolerances.
- Verification involves showing that a predictive model predicts adequately.
- A priori limits on performance must be identified in order to validate.
- Validation may be multivariate (multiple response variables).
- Desirability functions may help in validation or optimization.
- The general notions of equivalence and non-inferiority apply to validation.

Exercises and Questions

1. Consider a product having k continuously valued features, $x_1, x_2, \ldots, x_k$ and a single continuously valued key response variable, y. Suppose that using an approximating polynomial model, optimal values of the x_i were found. Describe a strategy for validating the product if the allowable tolerance for the optimal value of the response, y^*, is $\pm\delta$.
2. Are computations using predicted response values, where the prediction equation was generated using the design-generating data, sufficient for validating a product? Why or why not?
3. Consider a multivariate validation problem, where at least some of the response variables of interest do not have symmetric distributions. How important is this fact in determining acceptance limits for overall desirability?

Chapter 14
Simulation and Random Variable Generation

Originally, simulation meant using an electronic computer to generate pseudo-random numbers, uniformly distributed between 0 and 1 (Law and Kelton 1982). The single use of these numbers, called pseudo-random because they appear to have a uniform probability distribution, but in fact any sequence of them can be predicted exactly, was to perform numerical integration. Suppose an EAS wanted to compute a numerical approximation to an integral:

$$I = \int_a^b f(x)dx$$

where $f(x)$ is some fairly complicated and intractable function of x. Of course, nowadays there are many excellent numerical integration codes available. However, there was a time when such computing facilities were not easily obtained. It is true that if $f(U)$ is a function of a random variable, U, and U has density function $g(u)$, then

$$E[f(U)] = \int_{-\infty}^{+\infty} f(u)g(u)du$$

This fact is sometimes referred to as the "Law of the Unconscious Statistician" (Allen 2006). Now suppose U is uniformly distributed between a and b. Then

$$g(u) = \frac{1}{b-a} \; \forall u, \; a \leq u \leq b$$

And therefore:

$$I = (b-a)E(f(U)) = (b-a)\int_a^b f(u)\frac{1}{b-a}du$$

So, if we randomly generated N values of U, $u_1, u_2, \ldots, u_N$, and computed $y_1 = f(u_1), y_2 = f(u_2), \ldots, y_N = f(u_N)$ then an approximation to integral I is:

S.A. Pardo, *Empirical Modeling and Data Analysis for Engineers and Applied Scientists*,
DOI 10.1007/978-3-319-32768-6_14

$$\widehat{I} = (b-a)\frac{1}{N}\sum_{i=1}^{N} y_i$$

This integral approximation method was the original reason for what became known as Monte Carlo simulation. In Chap. 12, Monte Carlo simulation methods were used to assess Bayesian posterior distributions. It turns out that there can be some non-Bayesian applications of Monte Carlo simulation to the design problem.

Consider a deceptively simple-looking problem. An EAS wished to compute the middle 95 % of a force distribution. She knows that F = ma, and that the average mass is 5 g (0.005 kg) with a standard deviation of 0.0002 kg. She wants the average force to be approximately 1 N. In order to achieve a 1 N force, she believes she requires (on the average) an acceleration of 200 m/s^2. The question is how much standard deviation in both mass (m) and acceleration (a) should she allow so that the force will be between 1 N $\pm$ δ, with δ = 0.025 N, 95 % of the time? The variance of the product of two random variables is not the product of the variances. In fact, Springer (1979) has shown that the product of two normally distributed random variables is not necessarily normal. The analytic expressions for the product's density and cdf are in fact quite complicated, involving infinite series. Our EAS decides to try simulation. She uses the R code shown in Fig. 14.1, with N = 100,000 simulated values of mass and acceleration, and gets the histogram of forces shown in Fig. 14.2. The histogram appears quite symmetrically distributed. Figure 14.3 shows the mean, standard deviation 2.5th percentile, and 97.5th percentile of Force.

If Force were normally distributed, we would expect about 95 % of the results to fall between MForce $\pm$ 1.96*SForce = 0.9999829 $\pm$ 1.96(0.03997325) $\approx$ (0.9216, 1.0783). This is fairly close to the range of the 2.5th to 97.5th percentiles (0.9213, 1.0782). The standard deviations used for mass and acceleration were both 0.0002. The standard deviation of Force was approximately 0.0400. Certainly, the Force standard deviation is not the square root of the sums of variances of mass and acceleration. In order to help find distributional parameter values for mass and acceleration that will satisfy her requirements on Force, the EAS decides to use a full factorial experiment in four factors, with a center point run. Table 14.1 shows the factors and their low, high, and midpoint levels in natural units. Table 14.2 shows the experimental runs in Helmert-coded form. Table 14.3 shows the input values for mass and acceleration, and the Force statistics.

The results of the experiment were somewhat disappointing. The highest percent within the tolerance range (1 $\pm$ 0.025) was only 81.19 %, at the center point. The EAS suspects that the phenomenon has curvature, so she adds runs to make a Central Composite Design (CCD). The complete results are shown in Table 14.4.

Run 22 is promising. The middle 95 % was (0.9961, 1. 0039), and 100 % of all N = 100,000 simulated results fell within the tolerance range. This run had all factors set to their center level, except for the standard deviation of mass, set to its low level. Thus, through simulation the EAS has discovered that in order to meet the specifications for Force, she will need to reduce the variability in the Mass variable.

Although it seems that run 22 was best, the EAS wanted to know which factor had the greatest influence on the percent falling within the desired range for Force. Table 14.5 shows the coefficients of a full second-order model fit to the data with Helmert-coded regressors, sorted by model parameter estimate magnitudes. Note that the largest magnitudes were the squared terms. It appears that the greatest effect is average mass squared (MuM*MuM), followed by the standard deviation of mass squared (SigM*SigM). Without the simulation, and naively computing Force by multiplying average mass and average acceleration, the EAS would not have concluded that reducing mass variability was critical.

```
setwd("H:\\Personal Data\\Experimentation for Design & Validation\\Data &
Analyses\\")
#df1 <- read.csv("20141215 Example 13.2 F EQ MA.csv")
# Inputs:
# MuM - average mass (kg): L = 0.0045 H = 0.0055, C = 0.005
# SigM - std.dev. of mass (kg): L = 0.00001 H = 0.00018 C = 0.000095
# MuA - average acceleration (m/s**2): L = 195 H = 205 C = 200
# SigA - std.dev. of acceleration (m/s**2): L = 0.00001 H = 0.0002 C = 0.000105
# N = simulation sample size
# delta = tolerance for 95% range on Force
#
#attach(df1)
Force <- c()
Mass <- c()
Accel <- c()
intol <- c()

MuM <- 0.005
SigM <- 0.0002
MuA <- 200
SigA <- 0.0002
N <- 100000
delta <- 0.025
#
#
#
Mass <- rnorm(n = N, mean = MuM, sd = SigM)
Accel <- rnorm(n = N, mean = MuA, sd = SigA)
Force <- Mass*Accel
hForce <- hist(Force)
for (i in 1:N) {

 if ((Force[i] <= 1 + delta) & (Force[i] >= 1 - delta)) {
  intol[i] <- 1
  }
 else {
  intol[i] <- 0
  }
 }
```

Fig. 14.1 Force simulation R code

```
pintol <- mean(intol)
MForce <- mean(Force)
SForce <- sd(Force)
lq <- quantile(Force,probs=c(0.025))
uq <- quantile(Force,probs=c(0.975))
MForce
SForce
lq
uq
pintol
```

Fig. 14.1 (continued)

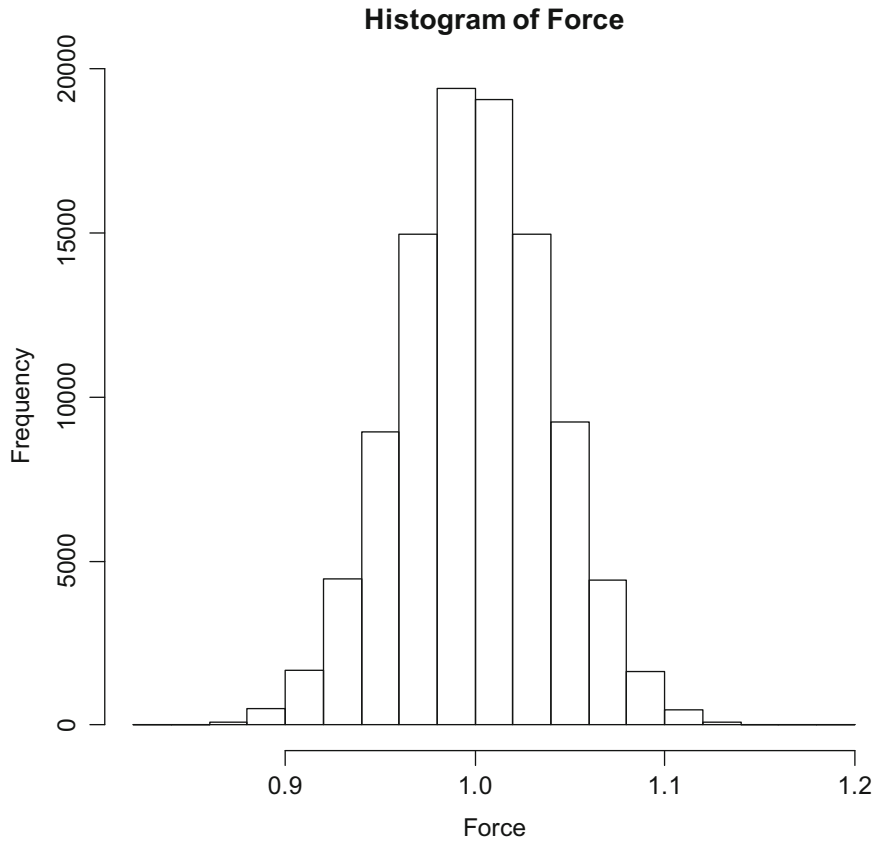


Fig. 14.2 Histogram of force, N = 100,000

The simulation has saved time and expense in creating multiple prototype systems. An actual experiment at the "optimal" conditions (average mass = 0.005 kg, standard deviation of mass = 0.00001 kg, average acceleration = 200 m/s^2, standard deviation of acceleration = 0.000105 m/s^2) must be performed, and Force must be measured. However, the search for good operational conditions is greatly reduced.

Fig. 14.3 Force statistics from simulation

```
> MForce
[1] 0.9999829
> SForce
[1] 0.03997325
> lq
     2.5%
0.9213199
> uq
   97.5%
1.078199
>pintol
0.46862
```

Table 14.1 Four factors and their levels

Factor	Low (−1)	High (+1)	Midpoint (0)	Units
MuM	0.0045	0.0055	0.005	kg
SigM	0.00001	0.00018	0.000095	kg
MuA	195	205	200	m/s^2
SigA	0.00001	0.0002	0.000105	m/s^2

Table 14.2 Full factorial experiment (Helmert-coded levels)

Pattern	MuM	SigM	MuA	SigA
−−−−	−1	−1	−1	−1
−−−+	−1	−1	−1	1
−−+−	−1	−1	1	−1
−−++	−1	−1	1	1
−+−−	−1	1	−1	−1
−+−+	−1	1	−1	1
−++−	−1	1	1	−1
−+++	−1	1	1	1
0	0	0	0	0
+−−−	1	−1	−1	−1
+−−+	1	−1	−1	1
+−+−	1	−1	1	−1
+−++	1	−1	1	1
++−−	1	1	−1	−1
++−+	1	1	−1	1
+++−	1	1	1	−1
++++	1	1	1	1

Another Example: Heat Transfer in a Bioreactor

An EAS wishes to design a thermal regulation system for a bioreactor that will be used to grow a tissue-engineered arterial graft. The bioreactor design is a cylinder of 2 mm thick crystal polystyrene tissue-culture plastic with an inner diameter of 5 mm. Cells will be seeded on the inner surface and liquid media (modeled as water) will be pumped through the length of the cylinder at 30 cm/s (approximate velocity of arterial blood). The bioreactor must be maintained at 37 ± 0.5 °C for optimal cell growth. The EAS decides to design a thermostat-based system that consists of a temperature sensor and a heater. The heat transfer through this system will be modeled as a combination of

Table 14.3 Input parameters and results for force simulation—full factorial

Run	Pattern	MuM	SigM	MuA	SigA	MForce	SForce	lq	uq	% In Tol.
1	−−−−	−1	−1	−1	−1	0.8775	0.00194	0.8737	0.8813	0.00 %
2	−−−+	−1	−1	−1	1	0.8775	0.00194	0.8737	0.8813	0.00 %
3	−−+ −	−1	−1	1	−1	0.9225	0.00204	0.9185	0.9265	0.00 %
4	−−++	−1	−1	1	1	0.9225	0.00205	0.9185	0.9265	0.00 %
5	−+−−	−1	1	−1	−1	0.8776	0.03512	0.8085	0.9461	0.26 %
6	−+−+	−1	1	−1	1	0.8774	0.03519	0.8084	0.9465	0.27 %
7	−++−	−1	1	1	−1	0.9224	0.03699	0.8498	0.9948	7.52 %
8	−+++	−1	1	1	1	0.9226	0.03684	0.8502	0.9951	7.47 %
9	0	0	0	0	0	1.0000	0.018985	0.9629	1.0371	81.19 %
10	+−−−	1	−1	−1	−1	1.0725	0.00195	1.0687	1.0763	0.00 %
11	+−−+	1	−1	−1	1	1.0725	0.00195	1.0687	1.0763	0.00 %
12	+−+−	1	−1	1	−1	1.1275	0.00205	1.1235	1.1315	0.00 %
13	+−++	1	−1	1	1	1.1275	0.00205	1.1235	1.1315	0.00 %
14	++−−	1	1	−1	−1	1.0725	0.03514	1.0039	1.1412	0.00 %
15	++−+	1	1	−1	1	1.0724	0.03497	1.0040	1.1410	8.52 %
16	+++−	1	1	1	−1	1.1275	0.036797	1.0557	1.2001	0.26 %
17	++++	1	1	1	1	1.1275	0.037044	1.0550	1.1999	0.27 %

Table 14.4 Central composite design for force simulation

Run	Pattern	MuM	SigM	MuA	SigA	MForce	SForce	lq	uq	% In Tol.
1	−−−−	−1	−1	−1	−1	0.8775	0.00194	0.8737	0.8813	0.00 %
2	−−−+	−1	−1	−1	1	0.8775	0.00194	0.8737	0.8813	0.00 %
3	−−+−	−1	−1	1	−1	0.9225	0.00204	0.9185	0.9265	0.00 %
4	−−++	−1	−1	1	1	0.9225	0.00205	0.9185	0.9265	0.00 %
5	−+−−	−1	1	−1	−1	0.8776	0.03512	0.8085	0.9461	0.26 %
6	−+−+	−1	1	−1	1	0.8774	0.03519	0.8084	0.9465	0.27 %
7	−++−	−1	1	1	−1	0.9224	0.03699	0.8498	0.9948	7.52 %
8	−+++	−1	1	1	1	0.9226	0.03684	0.8502	0.9951	7.47 %
9	0	0	0	0	0	1.0000	0.018985	0.9629	1.0371	81.19 %
10	+−−−	1	−1	−1	−1	1.0725	0.00195	1.0687	1.0763	0.00 %
11	+−−+	1	−1	−1	1	1.0725	0.00195	1.0687	1.0763	0.00 %
12	+−+−	1	−1	1	−1	1.1275	0.00205	1.1235	1.1315	0.00 %
13	+−++	1	−1	1	1	1.1275	0.00205	1.1235	1.1315	0.00 %
14	++−−	1	1	−1	−1	1.0725	0.03514	1.0039	1.1412	0.00 %
15	++−+	1	1	−1	1	1.0724	0.03497	1.0040	1.1410	8.52 %
16	+++−	1	1	1	−1	1.1275	0.036797	1.0557	1.2001	0.26 %
17	++++	1	1	1	1	1.1275	0.037044	1.0550	1.1999	0.27 %
18	000−	0	0	0	−1	1.0001	0.039949	0.9218	1.0785	46.97 %
19	000+	0	0	0	1	0.9998	0.04017	0.9216	1.0790	46.63 %
20	00−0	0	0	−1	0	0.9749	0.01857	0.9385	1.0112	49.50 %
21	00+0	0	0	1	0	1.0250	0.01948	0.9868	1.0632	49.79 %
22	**0−00**	**0**	**−1**	**0**	**0**	**1.0000**	**0.00200**	**0.9961**	**1.0039**	**100.00** %
23	0+00	0	1	0	0	0.9999	0.03605	0.9292	1.0706	51.40 %
24	−000	−1	0	0	0	0.9000	0.01898	0.8628	0.9374	0.00 %
25	+000	1	0	0	0	1.1001	0.01895	1.0632	1.1374	0.00 %

conduction through the polystyrene and convection due to the flow of media. The outer air temperature is assumed to be 25 °C.

The power output of the heater will increase or decrease depending on the value measured by the temperature sensor. The success of the thermal regulation system potentially depends on the noise

Table 14.5 Full second-order force model parameter estimates (sorted)

Term	Estimate	Abs(estimate)	Std error	t ratio	Prob > |t|
MuM	−0.00359	0.003594444	0.030901	−0.12	0.9097
MuA	0.003756	0.003755556	0.030901	0.12	0.9057
MuM*SigM	−0.00404	0.00404375	0.032775	−0.12	0.9043
SigM*MuA	0.004044	0.00404375	0.032775	0.12	0.9043
SigA	0.004528	0.004527778	0.030901	0.15	0.8864
SigM*SigA	0.005306	0.00530625	0.032775	0.16	0.8746
MuM*SigA	0.005356	0.00535625	0.032775	0.16	0.8734
MuA*SigA	−0.00536	0.00535625	0.032775	−0.16	0.8734
SigM	−0.01335	0.01335	0.030901	−0.43	0.6749
MuM*MuA	−0.01403	0.01403125	0.032775	−0.43	0.6776
MuA*MuA	**−0.08461**	**0.084614124**	**0.08215**	**−1.03**	**0.3273**
SigA*SigA	**−0.11306**	**0.113064124**	**0.08215**	**−1.38**	**0.1988**
SigM*SigM	**0.175936**	**0.175935876**	**0.08215**	**2.14**	**0.0579**
MuM*MuM	**−0.58106**	**0.581064124**	**0.08215**	**−7.07**	**<.0001**
Intercept	**0.614041**	**0.614040678**	**0.056607**	**10.85**	**<.0001**

Table 14.6 Heater factors and levels

Factor	Units	Low	High	Comment
sens.prec	Digits	2	4	Digits of numerical precision, sensor temp
sens.noise	Kelvin	3	5	Std.dev. of noise in sensor temperature measurement
sens.time	Seconds	3	5	Time delay for sensor reading
heat.prec	Digits	2	4	Digits of numerical precision, heater power output
heat.noise	Watts	3	5	Std.dev. of noise in heater power
heat.time	Seconds	9	11	Time delay for heater response
heat.factor	No units	0.001	0.002	Factor for incrementally increasing power proportional to temp

level and numerical precision of both the temperature sensor and the heating element. In addition, the EAS will consider the delay inherent in the temperature sensor or the heating element. Lower noise level, greater numerical precision, and shorter delay are expected to improve the thermal regulation, but also make the system components more expensive. Noise level is quantified as a standard deviation, and numerical precision is the number of decimal places to which temperature and heater power are computed. Through an experiment with the simulation program, the EAS hopes to determine the relative importance of the factors, and if possible, the minimum accuracy, minimum precision, and maximum delay required to maintain the bioreactor temperature within specification limits with a high degree of probability.

Table 14.6 shows all the factors under consideration, together with the range of values (low and high) the EAS would like to investigate.

With seven factors, the EAS decides to perform a ResIII fractional factorial, in order to estimate the relative importance of the factors. Figure 14.4 shows the Minitab alias structure output for the 2^{7-4} fractional design.

Figure 14.5 shows the runs with Helmert-coded factors.

This experimental design only requires eight runs. Since it is ResIII, it is most useful for ranking the importance of the factors, helping the EAS reduce the dimensions of the experimental factor space.

Figure 14.6 shows the R code that implements the simulation.

The temperature equation is based on heat conduction equation for a closed cylinder (Incropera and De Witt 1990):

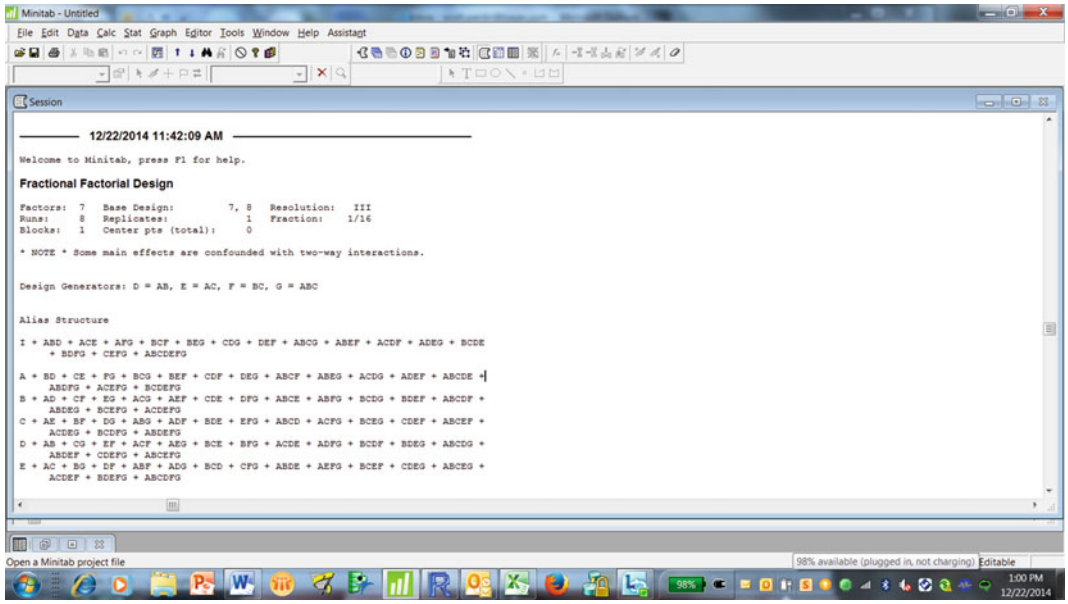
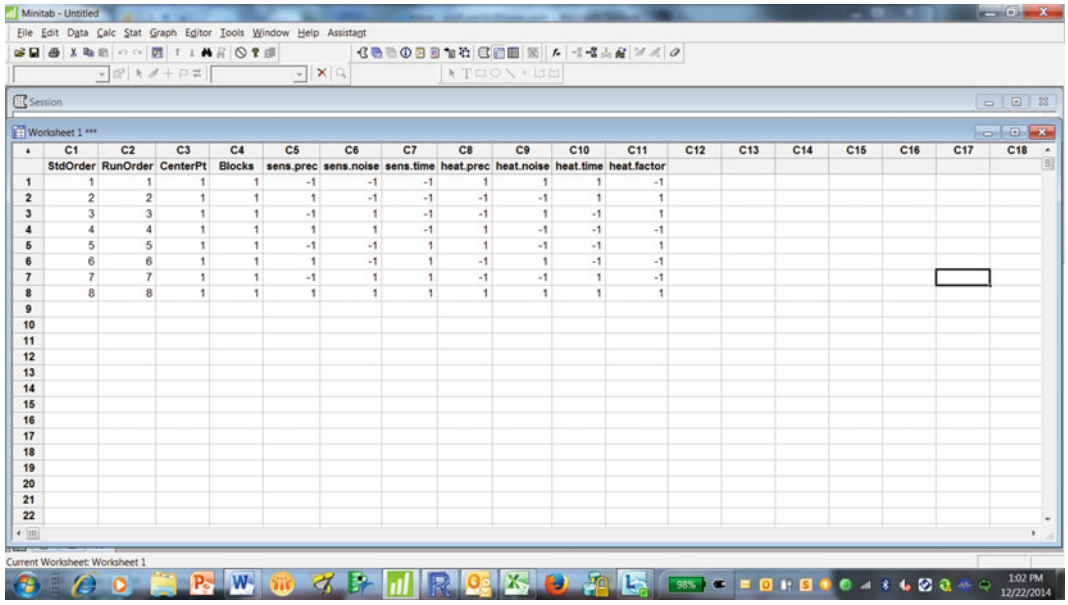


Fig. 14.4 Alias structure for 2^{7-4} design



	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18
	StdOrder	RunOrder	CenterPt	Blocks	sens.prec	sens.noise	sens.time	heat.prec	heat.noise	heat.time	heat.factor							
1	1	1	1	1	-1	-1	-1	1	1	1	-1							
2	2	2	1	1	1	-1	-1	-1	-1	1	1							
3	3	3	1	1	-1	1	-1	-1	1	-1	1							
4	4	4	1	1	1	1	-1	1	-1	-1	-1							
5	5	5	1	1	-1	-1	1	1	-1	-1	1							
6	6	6	1	1	1	-1	1	-1	1	-1	-1							
7	7	7	1	1	-1	1	1	-1	-1	1	-1							
8	8	8	1	1	1	1	1	1	1	1	1							
9																		
10																		
11																		
12																		
13																		
14																		
15																		
16																		
17																		
18																		
19																		
20																		
21																		
22																		

Fig. 14.5 Runs for the 2^{7-4} design

$$Q = 2\,k_{ps}\pi\,L\,\Delta T/\ln(r2/r1)$$

and on the convection equation:

$$Q = h\,A\,\Delta T.$$

```
setwd("H:\\Personal Data\\Experimentation for Design & Validation\\Data &
Analyses\\")
df1 <- read.csv("20141225 Example 14.2 Heat Transfer Inputs.csv")

sens.prec <- df1$sens.prec;
sens.noise <- df1$sens.noise;
sens.time <- df1$sens.time;
heat.prec <- df1$heat.prec;
heat.noise <- df1$heat.noise;
heat.time <- df1$heat.time;
heat.factor <- df1$heat.factor;

##############################
#Values for heat transfer calculations- All units SI (m, kg, etc.)
D <- 0.005; #inner diameter of bioreactor (characteristic length, meters)
L <- 0.1; #length, meters, of bioreactor
r1 <- 0.0025; #inner radius (meters)
r2<-0.0045; #outer radius (2 mm thickness)
kwater <- 0.58; # (W/(m*K))thermal conductivity of water- assumed to be fluid in
bioreactor
kps <- 0.13; #thermal conductivity of crystal polystyrene
Tair <- 298; #temperature of surrounding air (K)
A <- (pi*r1^2)*L; #surface area across which convection occurs
Nu <- 3.66; #Nusselt number for laminar internal flow in a cylinder assuming constant
wall temperature
h <- Nu*kwater/D; #convection coefficient

##############################
#####Initialize output vectors
mn<- c();
std<-c();
ave.prob.in.spec<-c();


#####
#Outer loop to run through simulations
for(k in 1:length(df1[,1])){

#######
#initialize loop variables
iterations <- 100; #number of simulation runs
totaltime <- 2100; #number of time units (sec) over which each simulation runs
mn.sum <- 0; #mean temperature
```

Fig. 14.6 Heat transfer simulation code

```
var.sum <- 0; #standard deviation of temperature
prob.in.spec <- c()

#simulation loop
for(i in 1:iterations){

        #initialize variables
        temp.set <- 37 + 273; #desired temperature of bioreactor
        temp.curr <- 25 + 273; #current temperature of bioreactor
        temp.dat <- c(); #stored values of temperature data
        smem <- rep(temp.curr,sens.time[k]); #memory of sensor
        hmem <- rep(0, heat.time[k]);#memory of heater
        Q <- 0; #initialize power
        temp.in.spec <- c();
        #runs each simulation for a specified period of time
        for(t in 1:totaltime){
                #update power output of heater
                Q <- Q + (temp.set-smem[1])*heat.factor[k];
                hmem <- hmem[-1];
                hmem <- append(hmem,round(Q + norm(1,0,heat.noise[k]),heat.prec[k]));

                #update actual temperature
                #reactor temperature in terms of power
                temp.curr <- Q/(2*kps*pi*L/(log(r2/r1))+ h*A) + Tair;
                temp.dat <- append(temp.dat, temp.curr);
                #update measured temperature
                smem <- smem[-1]
        #update sensor memory taking into account sensor noise and precision

        smem <- append(smem, round(temp.curr+rnorm(1, 0, sens.noise[k]),sens.prec[k]))

        if (temp.curr - 273 >= 36.5 & temp.curr - 273 <= 37.5) {
          temp.in.spec[t] <- 1
          }
        else {
          temp.in.spec[t] <- 0
          }
         }
         #collect summary statistics
         mn.sum <- mn.sum + mean((temp.dat-273)[300:length(temp.dat)])
         var.sum <- var.sum+var((temp.dat-273)[300:length(temp.dat)])
         prob.in.spec[i] <- mean(temp.in.spec[300:length(temp.dat)])
}

################################
```

Fig. 14.6 (continued)

```
#output
mn <- append(mn,mn.sum/iterations) #mean of means
std <- append(std,sqrt(var.sum/iterations))# sqrt of mean variance
# rmn <- round(mn,2)
# rstd <- round(std,2)
# time <- c(1:totaltime)/60 #generate vector for plotting x-axis in minutes
ave.prob.in.spec <- append(ave.prob.in.spec,mean(prob.in.spec));

}
df2 <- cbind(df1,mn,std,ave.prob.in.spec)
write.csv(df2,file="20141225 Example 13.2 Heat Transfer Outputs.csv")
```

Fig. 14.6 (continued)

Table 14.7 The 2^{7-4} + center point runs with results

Run	sens. prec	sens. noise	sens. time	heat. prec	heat. noise	heat. time	heat. factor	mn	std	ave.prob.in. spec
1	2	3	3	4	5	11	0.0010	36.901	0.2956	0.9136
2	4	3	3	2	3	11	0.0010	36.907	0.2948	0.9163
3	2	5	3	2	5	9	0.0010	36.908	0.3615	0.8392
4	4	5	3	4	3	9	0.0010	36.885	0.3654	0.8294
5	2	3	5	4	3	9	0.0020	36.988	0.2549	0.9405
6	4	3	5	2	5	9	0.0020	37.001	0.2493	0.9470
7	2	5	5	2	3	11	0.0020	37.005	0.4302	0.7362
8	4	5	5	4	5	11	0.0020	36.982	0.4244	0.7387
9	3	4	4	3	4	10	0.0015	36.969	0.2946	0.8892
9	3	4	4	3	4	10	0.0015	36.967	0.2858	0.8975
9	3	4	4	3	4	10	0.0015	36.999	0.2867	0.9045

The variables and constants are:

k_{ps} = thermal conductivity of polystyrene ≈ 0.13

L = length

r2 = outer radius

r1 = inner radius

$h = Nu*k_{water}/D$ = heat transfer coefficient (convection)

Nu = Nusselt number for laminar internal flow in a cylinder assuming constant wall temperature ≈ 3.66

k_{water} = thermal conductivity constant for water ≈ 0.58

D = inner diameter of bioreactor

ΔT = Current temperature inside bioreactor—initial temperature (i.e., air temperature)

The EAS decides that in addition to the eight runs, he will replicate at a center point run (run 9). Table 14.7 shows the runs and the results.

The main effects model was fit to the average proportion of temperatures within the 37.0 ± 0.5 °C specifications. This response was called "ave.prob.in.spec" in the simulation program. The model was fit using Helmert-coded factor levels. The results are given in Fig. 14.7. Note that the residual plot indicates the presence of second-order effects, as evidenced by the larger residual values at the center point. The only factor that was significant was sens.noise. Using the sorted parameter estimate table, and given the indication of the presence of second-order effects, the EAS decides to perform a

Response ave.prob.in.spec

Summary of Fit

RSquare	0.93608
RSquare Adj	0.786934
Root Mean Square Error	0.034223
Mean of Response	0.868373
Observations (or Sum Wgts)	11

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	7	0.05145543	0.007351	6.2763
Error	3	0.00351361	0.001171	**Prob > F**
C. Total	10	0.05496904		0.0797

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	1	0.00339629	0.003396	57.8945
Pure Error	2	0.00011733	0.000059	**Prob > F**
Total Error	3	0.00351361		0.0168*
				Max RSq

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.8683727	0.010319	84.16	<.0001*
sens.prec	0.0002375	0.0121	0.02	0.9856
sens.noise	-0.071738	0.0121	-5.93	0.0096*
sens.time	-0.017013	0.0121	-1.41	0.2544
heat.prec	-0.002062	0.0121	-0.17	0.8755
heat.noise	0.0020125	0.0121	0.17	0.8785
heat.time	-0.031413	0.0121	-2.60	0.0806
heat.factor	0.0010625	0.0121	0.09	0.9356

Residual by Predicted Plot

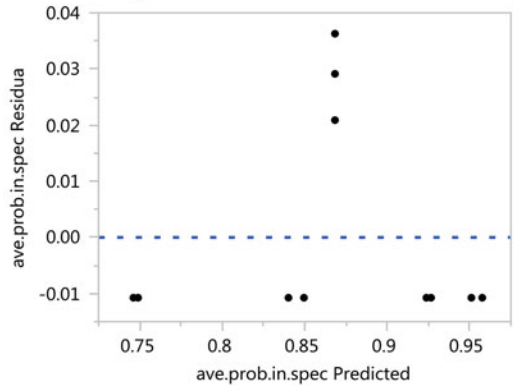


Fig. 14.7 Regression fit (from JMP) for main effects model

Sorted Parameter Estimates

Term	Estimate	Std Error	t Ratio	t Ratio	Prob>|t|
sens.noise	-0.071738	0.0121	-5.93		0.0096*
heat.time	-0.031413	0.0121	-2.60		0.0806
sens.time	-0.017013	0.0121	-1.41		0.2544
heat.prec	-0.002062	0.0121	-0.17		0.8755
heat.noise	0.0020125	0.0121	0.17		0.8785
heat.factor	0.0010625	0.0121	0.09		0.9356
sens.prec	0.0002375	0.0121	0.02		0.9856

Fig. 14.7 (continued)

Table 14.8 Levels for the Box–Behnken experiment

Factor	Units	Low	High	Comment
sens.prec	Digits		4	Digits of numerical precision, sensor temp
sens.noise	Kelvin	2	4	Std.dev. of noise in sensor temperature measurement
sens.time	Seconds	4	6	Time delay for sensor reading
heat.prec	Digits	2		Digits of numerical precision, heater power output
heat.noise	Watts	1		Std.dev. of noise in heater power
heat.time	Seconds	9	11	Time delay for heater response
heat.factor	No units		0.002	Factor for incrementally increasing power proportional to temp

Table 14.9 Runs with coded levels for the Box–Behnken experiment

Patte rn	sens.noise	heat.time	sens.time
−−0	−1	−1	0
−0−	−1	0	−1
−0+	−1	0	1
−+0	−1	1	0
0−−	0	−1	−1
0−+	0	−1	1
0	0	0	0
0	0	0	0
0	0	0	0
0+−	0	1	−1
0++	0	1	1
+−0	1	−1	0
+0−	1	0	−1
+0+	1	0	1
0	1	1	0

second-order experiment in the top three factors, namely sens.noise, heat.time, and sens.time. Furthermore, he does some exploratory runs and decides to alter the ranges of the factors. The EAS chooses a Box–Behnken experiment with three replicates at the center point. The other factors were kept constant. Table 14.8 shows the factor levels for the second-order experiment.

Table 14.9 shows the coded levels of the three factors varied in the Box–Behnken experiment. Table 14.10 shows the natural levels, together with the fixed values of the other factors not included, and the data.

Figure 14.8 shows the second-order model fit to the ave.prob.in.spec response.

The only significant effect was sens.noise*sens.noise. A reduced second-order model was then fit, with only the sens.noise, heat.time, sense.noise*heat.time (not significant in the full model, but having a coefficient twice as big as the next largest coefficient), and sens.noise*sens.noise terms included. Figure 14.9 shows the results of the fit.

Table 14.10 Box–Behnken runs with natural levels, fixed values, and data

Run	sens. prec	sens. noise	sens. time	heat. prec	heat. noise	heat. time	heat. factor	mn	std	ave.prob.in. spec
1	4	2	5	2	1	9	0.002	37.006	0.1687	0.9968
2	4	2	4	2	1	10	0.002	37.008	0.1645	0.9970
3	4	2	6	2	1	10	0.002	36.997	0.1680	0.9968
4	4	2	5	2	1	11	0.002	36.999	0.1688	0.9963
5	4	3	4	2	1	9	0.002	36.994	0.2469	0.9463
6	4	3	6	2	1	9	0.002	36.995	0.2520	0.9433
7	4	3	5	2	1	10	0.002	36.996	0.2521	0.9441
7	4	3	5	2	1	10	0.002	36.998	0.2579	0.9408
7	4	3	5	2	1	10	0.002	36.980	0.2604	0.9347
8	4	3	4	2	1	11	0.002	36.979	0.2555	0.9404
9	4	3	6	2	1	11	0.002	36.997	0.2562	0.9411
10	4	4	5	2	1	9	0.002	36.998	0.3439	0.8409
11	4	4	4	2	1	10	0.002	36.993	0.3347	0.8479
12	4	4	6	2	1	10	0.002	36.975	0.3366	0.8466
13	4	4	5	2	1	11	0.002	36.997	0.3304	0.8581

Now all the terms, except for the main effect of heat.time, in the model have significant coefficients. However, since the sens.noise*heat.time term was significant, the heat.time main effect will be left in the model.

Figure 14.10 shows the Optimization for the reduced second-order model, using the JMP Custom Profiler function. The optimal design has sens.noise $= -1$ (2 K) and heat.time $= -1$ (9 s). These values are predicted to virtually insure with 100 % probability that the bioreactor will remain in the temperature range 37 $\pm$ 0.5 °C.

One note of information about sample means taken over time series data is appropriate. The mean of the temperatures within a simulation replicate is calculated over temperatures observed in a time sequence, or time series. Thus, the values of temperature are not independent of each other. In the case of a random sample, the sample average has estimated variance:

$$\widehat{V}(\bar{x}) = \frac{s^2}{n}$$

In the case of an average of time series data (which are said to be autocorrelated, or correlated with previous values in time), the estimated variance is (Cryer 1986):

$$\widehat{V}(\bar{x}) = \frac{s^2}{n}\left[1 + 2\sum_{k=1}^{n-1}\left[1 - \frac{k}{n}\right]\widehat{\rho}_k\right]$$

where:

$$\widehat{\rho}_k = \frac{\sum_{j=1}^{n-k}\left(x_j - \bar{x}\right)\left(x_{j+k} - \bar{x}\right)}{\sum_{j=1}^{n}\left(x_j - \bar{x}\right)^2}$$

and:

Response ave.prob.in.spec

Summary of Fit

RSquare	0.997248
RSquare Adj	0.992295
Root Mean Square Error	0.00501
Mean of Response	0.93141
Observations (or Sum Wgts)	15

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	9	0.04547442	0.005053	201.3282
Error	5	0.00012548	0.000025	**Prob > F**
C. Total	14	0.04559991		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.9398964	0.002892	324.96	<.0001*
sens.noise	-0.074176	0.001771	-41.88	<.0001*
heat.time	0.0010737	0.001771	0.61	0.5708
sens.time	-0.000494	0.001771	-0.28	0.7914
sens.noise*heat.time	0.0044434	0.002505	1.77	0.1363
sens.noise*sens.time	-0.00028	0.002505	-0.11	0.9152
heat.time*sens.time	0.0009078	0.002505	0.36	0.7318
sens.noise*sens.noise	-0.018795	0.002607	-7.21	0.0008*
heat.time*heat.time	0.0019015	0.002607	0.73	0.4985
sens.time*sens.time	0.0009812	0.002607	0.38	0.7221

Residual by Predicted Plot

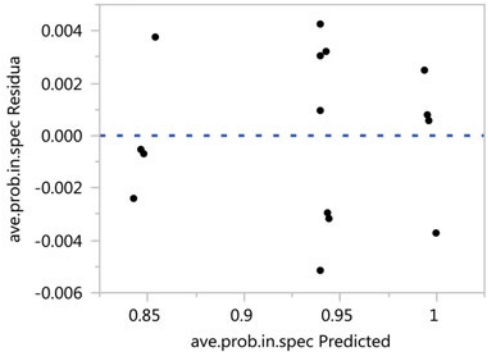


Fig. 14.8 JMP output—full second-order model Fit

Sorted Parameter Estimates

Term	Estimate	Std Error	t Ratio	t Ratio	Prob>|t|
sens.noise	-0.074176	0.001771	-41.88		<.0001*
sens.noise*sens.noise	-0.018795	0.002607	-7.21		0.0008*
sens.noise*heat.time	0.0044434	0.002505	1.77		0.1363
heat.time*heat.time	0.0019015	0.002607	0.73		0.4985
heat.time	0.0010737	0.001771	0.61		0.5708
sens.time*sens.time	0.0009812	0.002607	0.38		0.7221
heat.time*sens.time	0.0009078	0.002505	0.36		0.7318
sens.time	-0.000494	0.001771	-0.28		0.7914
sens.noise*sens.time	-0.00028	0.002505	-0.11		0.9152

Fig. 14.8 (continued)

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Note that the sample variance estimator used here has the factor $\frac{1}{n}$ instead of the more usual $\frac{1}{n-1}$. This is more for convenience of computation. Practically speaking, the use of n versus $n-1$ usually has little impact.

The function, $\widehat{\rho}_k$, is the sample autocorrelation function (acf). The values of the variable k are called "lags". The R function acf(x) produces the sample autocorrelation function. As an example, consider the time series generated by the heat transfer simulation, stored in temp.dat. The R command for computing the acf for the steady state temperature data (beginning 5 min after the simulation began) is:

rho <- acf(temp.dat[300:2100], lag.max = 1799)

Figure 14.11 shows the plot of the acf which is automatically created by the R function.

The new vector, rho, stores the acf values.

The square root of $\widehat{V}(\overline{x})$ is the estimated standard error of the mean, and is required for computing confidence intervals for the mean.

In this chapter, two examples of how simulation involving the generation of random variables can assist in making design decisions. Both of these examples incorporated prior knowledge of the physical aspects of the systems under consideration, but in a non-Bayesian manner. Although equations that describe the phenomena were known, when uncertainty about input parameters or when uncontrollable sources of noise exist, simulation aids in predicting and optimizing system response.

Key Points

- Monte Carlo simulation involves the generation of random variable values using a computer program.
- Even simple equations of physics cannot necessarily be used to determine the distributional characteristics of a response variable when the inputs are assumed to be random variables.
- Once input parameters are determined, a simulation program can be constructed, and designed experiments with the program can be performed to characterize the response distribution.
- Experimenting with a simulation program is less expensive than creating multiple prototypes.
- Simulation models may be non-Bayesian in nature.
- Any conclusions drawn from a simulation must be verified by actual experimentation.

Response ave.prob.in.spec

Summary of Fit

RSquare	0.996777
RSquare Adj	0.995487
Root Mean Square Error	0.003834
Mean of Response	0.93141
Observations (or Sum Wgts)	15

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	0.04545292	0.011363	773.0695
Error	10	0.00014699	0.000015	**Prob > F**
C. Total	14	0.04559991		<.0001*

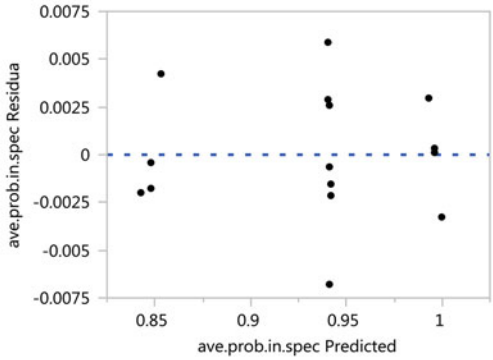
Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	4	0.00009590	0.000024	2.8159
Pure Error	6	0.00005109	8.514e-6	**Prob > F**
Total Error	10	0.00014699		0.1242
				Max RSq

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.9415436	0.001449	649.75	<.0001*
sens.noise	-0.074176	0.001355	-54.72	<.0001*
heat.time	0.0010737	0.001355	0.79	0.4467
sens.noise*heat.time	0.0044434	0.001917	2.32	0.0429*
sens.noise*sens.noise	-0.019001	0.001984	-9.58	<.0001*

Residual by Predicted Plot



Sorted Parameter Estimates

Term	Estimate	Std Error	t Ratio	t Ratio	Prob>|t|
sens.noise	-0.074176	0.001355	-54.72		<.0001*
sens.noise*sens.noise	-0.019001	0.001984	-9.58		<.0001*
sens.noise*heat.time	0.0044434	0.001917	2.32		0.0429*
heat.time	0.0010737	0.001355	0.79		0.4467

Fig. 14.9 Reduced model second-order fit

Factor	Current X	Lock	Lo Limit	Hi Limit
sens.noise	-1	☐	-1	1
heat.time	-1	☐	-1	1

Response	Current Y	Lo Limit	Hi Limit	Benchmark
Pred Formula ave.prob.in.spec	1.0000881	.	.	0.9415436

Optimization

Objective	Trips	Max Cycles	Max Iter	Convergence Limit	Convergence Criterion
1.0000881453	20	50	250	0.000001	0

Maximize

Fig. 14.10 Optimization of reduced second-order model—from JMP custom profiler

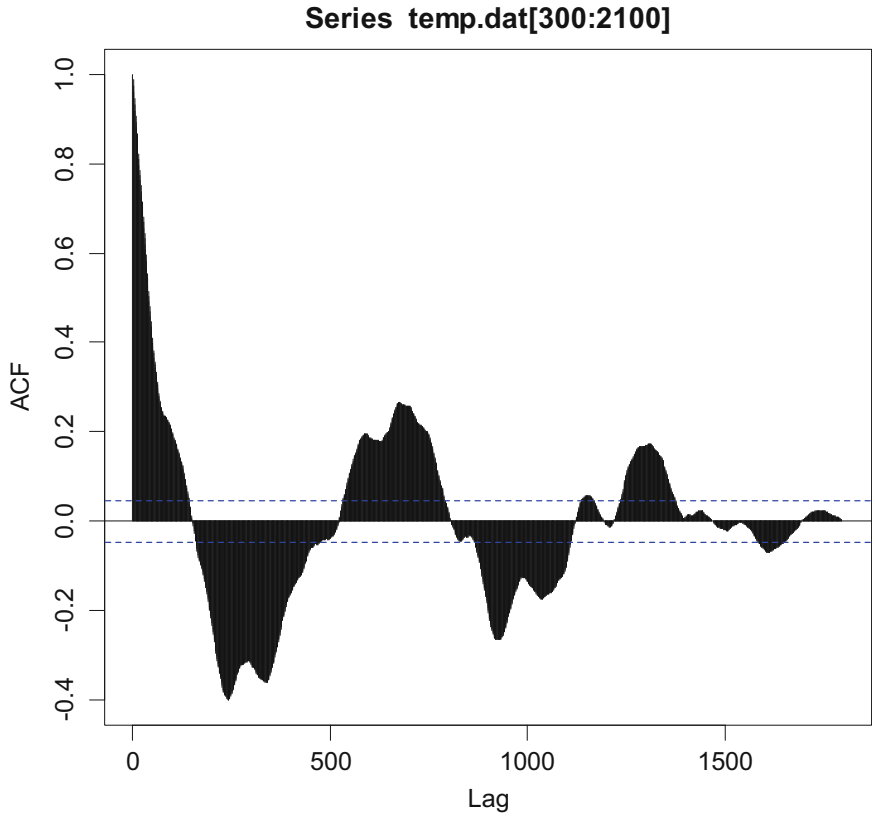


Fig. 14.11 Plot of the time.dat ACF

Exercises and Questions

1. Consider the Ideal Gas Law: PV = nRT, where P is pressure, V is volume (of a gas in a closed container), n is the moles of gas, R is the gas' constant (i.e., a constant specific to the type of gas), and T is temperature. Suppose that V and T vary probabilistically, having normal distributions with some range of values for their respective means and standard deviations. Design an experiment with a simulation to assess the distribution of P, and how it is affected by the means and standard deviations of V and T, and by the value of nR.
2. In the Force example of the text, what, if anything, would you do differently? Would you have changed the range of factor levels after the first experimental run?
3. Based on the results in Table 14.4, would you guess that the normal distribution is a reasonable approximation to the distribution of Force? Hint: consider the interval mean ± 1.96*standard deviation of Force, and the lower and upper quantiles of the simulated Force distribution at optimal conditions (Table 14.4, Run #22). You may also use a Shapiro–Wilk or other test for normality.
4. Fit a first-order main effects model to the response "mn" in Table 14.7. Do you conclude the same thing about the presence of second-order effects as in the case of the ave.prob.in.spec response?
5. Fit a second-order model to the "mn" response using the data in Table 14.10. Use Helmert-coded factors. Do you conclude the same thing about the significance of factors as in the fit for ave.prob.in.spec?

Chapter 15
Taguchi Methods® and Robust Design

Dr. Genichi Taguchi (Taguchi and Wu 1980) described methods for using designed experiments in designing and improving products and processes. The Taguchi approach differs from the methods described in previous chapters, mostly due to their lack of fitting polynomial models to data. Nevertheless, Taguchi methods can be a useful tool in developing product or process designs. We will describe some of the rudiments of his methods, together with some additional ideas. The elements we will consider are the quadratic loss function and signal-to-noise ratios, control and noise parameters, and the designs referred to as orthogonal arrays.

The Quadratic Loss Function

Presume for the moment that the variable X represents a measure of quality. Furthermore, suppose that ideally $X = \tau$, which we will call the target value for X.

The "loss" is the degree to which X differs from τ. For reasons that may not be apparent, we will measure this loss in the squared deviation:

$$L(X) = k(X - \tau)^2$$

The value of k only determines the rate at which the loss changes, so we will generally let $k = 1$.

The objective for robust design is to find values of control parameters that minimize the average loss in the face of noise parameters whose values vary in an uncontrolled fashion (except in the experimentation process).

The loss function can be applied to situations where it is desirable to minimize X or maximize X. Let $\tau = 0$. The, the loss function is:

$$L(X) = kX^2$$

Clearly minimizing X will minimize the loss, so that the objective of minimizing loss is achieved by minimizing X. Similarly, let:

$$Y = \frac{1}{X}$$

S.A. Pardo, *Empirical Modeling and Data Analysis for Engineers and Applied Scientists*,
DOI 10.1007/978-3-319-32768-6_15

so that the loss with respect to Y is:

$$L(Y) = kY^2 = k\frac{1}{X^2}$$

Thus minimizing the loss is achieved by maximizing X.

One might ask why, especially in the case of minimizing or maximizing X, it is valuable to use the quadratic loss function. The answer lies in the expected value of $L(X)$. For our discussions, we will simply let $k = 1$. Then if $\mathrm{E}[X] = \mu$ is the expected value of X, and $\mathrm{V}[X] = \sigma^2$ is the variance of X, then it can be derived that:

$$E[L(X)] = E\left[(X-\tau)^2\right] = (\mu-\tau)^2 + \sigma^2$$

Thus, minimizing the average quadratic loss simultaneously minimizes the average difference of X from target and the variability of X. When one statistic is simultaneously related to mean and variance, so that either having a mean closer to target or reducing variability gives it a more desirable value, it has been called a "concurrent" statistic (Barker 1990).

Some people may have an aversion to minimizing, and Dr. Taguchi recognized this. So, he created what he called the "signal-to-noise" ratio (*SNR*). The *SNR* is nothing more than the loss function expressed in decibels. So, minimizing the expected value of the loss function is equivalent to maximizing the expected value of:

$$SNR = -10log_{10}L(X)$$

Aside from an emotional tie to maximization, there is no particular reason to use the *SNR* transformation of the loss function. Taguchi's original work did not concern itself with the statistical significance of parameters; he presumed that all parameters were meaningful, and that their effects were already known to be repeatable. However, it may be valuable to be able to assess the significance of design parameters. As such, it will be much easier to use the loss function directly, as its statistical sample properties are easier to derive than are those of the *SNR*.

Suppose a sample of values of X is obtained, $x_1, x_2, \ldots, x_n$. Then the sample average loss function is:

$$\widehat{L}(x_1, x_2, \cdots, x_n) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\tau)^2$$

If the variable X is normally distributed with expected value μ and variance σ^2, then:

$$\widehat{L} = \frac{\sigma^2}{n}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}+\frac{\mu-\tau}{\sigma}\right)^2 \sim \frac{\sigma^2}{n}\chi'^2\left(n,\ \lambda = \frac{n(\mu-\tau)^2}{\sigma^2}\right)$$

$\chi'^2\left(n,\ \lambda = \frac{n(\mu-\tau)^2}{\sigma^2}\right)$ stands for a non-central Chi-squared random variable with n degrees of freedom and *non-centrality parameter* λ (Johnson et al. 1995). Although both μ and σ^2 are generally unknown, an approximate $100(1-2\alpha)$ % confidence interval is

$$\left(\frac{\widehat{\sigma}^2}{n}\chi_{\alpha}^{'2}\left(n,\,\widehat{\lambda}=n\widehat{Z}_{\tau}^2\right),\;\frac{\widehat{\sigma}^2}{n}\chi_{1-\alpha}^{'2}\left(n,\,\widehat{\lambda}=n\widehat{Z}_{\tau}^2\right)\right)$$

where:

$$\widehat{Z}_{\tau}=\frac{\tau-\widehat{\mu}}{\widehat{\sigma}}$$

and $\chi_p^{'2}\left(n,\widehat{\lambda}\right)$ = the 100pth percentile of a non-central Chi-squared distribution with n degrees of freedom and non-centrality parameter $\widehat{\lambda}$.

The statistics $\widehat{\mu}$, $\widehat{\sigma}$ are sample estimates of μ and σ, respectively. Since the loss function is in squared units, it is somewhat uninterpretable as it stands. One possible remedy is to divide it by τ^2, so that the loss is expressed as a proportion (or percent) of the target value, and take its square root (and the square roots of the confidence limits divided by τ^2). We will call this computation the root mean square percent loss.

Suppose we wanted to compare average loss between two different designs for a product. If the null hypothesis is that the average loss does not differ with respect to variable X, then if $\widehat{L}_1$ represents the average loss for a sample of n_1 observations of X with design 1 and $\widehat{L}_2$ the average loss for a sample of n_2 observations of X with design 2, then the ratio:

$$\widehat{F}=\frac{\frac{n_1}{\sigma^2}\frac{\widehat{L}_1}{n_1}}{\frac{n_2}{\sigma^2}\frac{\widehat{L}_2}{n_2}}=\frac{\widehat{L}_1}{\widehat{L}_2}$$

has a doubly non-central F distribution (F'') with n_1 and n_2 numerator and denominator degrees of freedom, respectively and numerator and denominator non-centrality parameters:

$$\lambda_k=\frac{n_k(\mu-\tau)^2}{\sigma^2},\quad k=1,\,2$$

respectively. Now we have a means of deciding whether to believe that the average loss for one design is different from that of another. Compare the sample statistic $\widehat{F}$ to the lower and upper tails of a doubly non-central F with n_1 and n_2 degrees of freedom and non-centrality parameters λ_k, $k = 1$, 2. The only problem is that the parameters μ and σ are unknown. If we are also hypothesizing that $\mu = \tau$, then the non-centrality goes away, so we are only comparing the statistic to percentiles of the usual (central) F distribution. If, however, we are not willing to make that assumption, then one possibility is to aggregate all the data from both designs and compute the sample average and standard deviation, call them $\widehat{\mu}$ and $\widehat{\sigma}$. These can then be used to "approximate" the actual numerator and denominator non-centralities. Thus the estimated non-centrality parameters for the numerator and denominator would be:

$$\widehat{\lambda}_k=\frac{n_k(\widehat{\mu}-\tau)^2}{\widehat{\sigma}^2},\quad k=1,\,2$$

Inasmuch as percentiles of doubly non-central F distributions are somewhat hard to find, a reasonable approximation may be obtained with the formula described by Johnson et al. (1995). That is, if $F''(p,\,n_1,\,n_2,\,\lambda_1,\,\lambda_2)$ represents a percentile of a doubly non-central F, and $F(p,\,n_1,\,n_2,\,\lambda_1)$ the same percentile for a singly non-central F (with non-centrality λ_1 in the numerator only) then:

$$F'' = \left(1 + \frac{\lambda_2}{n_2}\right)^{-1} F'$$

Figure 15.1 shows some R code for computing the approximate percentiles of a doubly non-central F.

```
setwd("H:\\Personal Data\\Experimentation for Design & Validation\\Data &
Analyses\\")
df1 <- read.csv("20150325 non-central F parameters.csv")
#
#
attach(df1)

# Ref: Johnson, Kotz, Balakrishnan, (1995) Continuous Univariate Distributions
#                              Vol. 2, 2nd Ed. p. 502 John Wiley and Sons, New York


#
# inputs:
# nu1 = sample size 1
# nu2 = sample size 2
# Target
# mu = hypothetically common mean
# sigma = hypothetically common standard deviation
#
#
alpha <- 0.025
Ztau <- (Target - mu) / sigma

lamda1 <- nu1*(Ztau**2)
lamda2 <- nu2*(Ztau**2)

Fprime_low <- qf(alpha,df1=nu1,df2=nu2,ncp=lamda1) # this is the lower tail value for
#                                                    singly noncentral F
Fprime_high <- qf(1-alpha,df1=nu1,df2=nu2,ncp=lamda1) # this is the upper tail value
#                                                       for singly non-central F

xfactor <- 1 + (lamda2/nu2)

Fdp_low <- Fprime_low / xfactor; # lower tail value doubly non-central F
Fdp_high <- Fprime_high / xfactor; #upper tail value doubly non-central F

xfactor
Fprime_low
Fprime_high
Fdp_low
Fdp_high
```

Fig. 15.1 Computing percentiles of doubly non-central F distributions

For inputs:

nu1 = nu2 = 8, Target = 100, mu = 95, and sigma = 1.2, the outputs are:

```
> xfactor
[1] 18.36111
> Fprime_low
[1] 8.898236
> Fprime_high
[1] 51.24304
> Fdp_low
[1] 0.4846241
> Fdp_high
[1] 2.790846
```

Fig. 15.1 (continued)

Parameter Design: Noise Parameters, Control Parameters, Inner and Outer Arrays

Taguchi identified the need for choosing designs that would be about to perform under the normally uncontrollable conditions of the environment in which they would be used. The optimization step is referred to as "Parameter Design". He divided the design parameters into two groups:

Control: those things whose values the EAS can choose; Noise: those things whose values that the EAS cannot choose, but must have his or her design perform well regardless of those values

For example, an engineer may choose the materials and dimensions of a cardiac lead, the cable that delivers electrotherapy from a pacemaker to the heart, but she cannot choose the impedance that the heart tissue will present. The materials and dimensions are control parameters, whereas the impedance is a noise parameter.

Dr. Taguchi suggested that two experimental designs should be performed together. The designs are referred to as the inner array (for control parameters) and the outer array (for noise parameters). The idea is that for every run in the inner array, a complete experiment in the noise parameters (outer array) should be performed. The noise array would be constructed by varying in a controlled fashion the noise parameters, in order to have a measure of loss over a wider range of noise "conditions" for each combination of the control parameters. In that way, the combination of the control parameter values that minimizes the loss (or maximizes the SNR) in face of the noise parameter variation would be the most "robust" choice for the product or process design.

The experimental designs employed by Taguchi are of a general class called "Orthogonal Arrays". All fractional factorials (or, for that matter, full factorials) are orthogonal arrays, but not all orthogonal arrays are fractional factorials. In order to simplify the choices of experimental design, Taguchi created a sort of catalogue of orthogonal arrays, categorized by the number of runs (rows). He used the letter "L", which stands for Latin Square, together with the number of rows to designate the design. A Latin Square is a $k \times k$ array of k unique items, usually symbolized with capital Latin letters, such that each item appears in each column and each row exactly once. The orthogonal array is a sort of generalization of the Latin Square. Thus, the "L8" array has eight rows, or runs, and it can be used with up to seven parameters (or factors, as we have called them). The fewer the number of factors (columns) used from a given array, the greater the resolution. Figure 15.2 shows the L8 array, as generated using JMP 11. This is a two-level design, with the "low" level designated with a "1" and the "high" level with a "2". That is, Taguchi preferred to represent the levels as positive integers,

Fig. 15.2 The Taguchi L8 array

X1	X2	X3	X4	X5	X6	X7	Pattern
1	1	1	1	1	1	1	−−−−−−−
1	1	1	2	2	2	2	−−−++++
1	2	2	1	1	2	2	−++−−++
1	2	2	2	2	1	1	−++++−−
2	1	2	1	2	1	2	+−+−+−+
2	1	2	2	1	2	1	+−++−+−
2	2	1	1	2	2	1	++−−++−
2	2	1	2	1	1	2	++−+−−+

Fig. 15.3 The L4 array

N1	N2	Pattern
1	1	−−
1	2	−+
2	1	+−
2	2	++

X1	X2		X4				N1=1,N2=1	N1=1,N2=2	N1=2,N2=1	N1=2,N2=2
1	1		1							
1	1		2							
1	2		1							
1	2		2							
2	1		1							
2	1		2							
2	2		1							
2	2		2							

Fig. 15.4 Combined L8 and L4 arrays

rather than with Helmert coding. With $k = 7$ parameters, this is a 2^{7-4} fractional factorial design of Resolution III. If only columns X1, X2, and X4 were used (i.e., only $k = 3$ parameters) this array would constitute a 2^3 full factorial design.

The L4 array is given in Fig. 15.3. The L4 is a 2^2 full factorial design in two factors. Note that here the columns have been designated as N1 and N2, and we will use this to illustrate the noise array.

Suppose the EAS has decided that there are $k = 3$ control parameters, and k = 2 noise parameters. Then he would combine the L8 and L4 arrays, so that each run in the L8array has all four of the L4 array completed. Figure 15.4 illustrates the combined arrays.

The empty cells in the Figure represent the quadratic loss values, given target τ. The columns not being used are blacked out. The first row represents a run where all three control parameters are set to their "low" level. The eighth row is a run where all three control parameters are set to their "high" level. For each row, observations are obtained at four different "noise" conditions.

The idea would be to average the loss over all rows where a given parameter is set to its "low" value, and compare it to the loss averaged over all the rows where the parameter was set to its "high level". The parameter value to use in the design would be the one that yielded the lowest average loss. This methodology is a considerable departure from all the methods discussed so far, in that we are not finding an approximating polynomial to predict optimal factor (or regressor, or in the Taguchi

terminology, parameter) values or settings. The biggest disadvantage is that the Taguchi approach does not allow us to interpolate between levels. The advantage is that it is very simple to implement.

Example: Pharmaceutical Tablet Dissolution

The EAS is designing a tablet for delivering a drug orally. It is desired for the tablet to dissolve at a certain rate. In particular, the measure of goodness is to have half the tablet's mass dissolved in the stomach after 4 h. The time for 50 % of the tablet to dissolve will be called D_{50}. The target value for D_{50} is 240 min = 4 h. There are three control parameters of interest in this case:

T = processing temperature of the tablet formation (°C);
C = drying and time (minutes)
P = force used to compact the wet ingredients into the tablet form (Newtons).
The two noise parameters are:
pH = The pH of stomach contents;
V = volume of stomach contents prior to introducing the tablet.

An experimental chamber, in which both pH and V can be controlled, will be used to perform the tests. The L8 and L4 arrays will be employed. The levels for the L8 (control parameters) are given in Table 15.1, and the levels for the L4 array are in Table 15.2. The columns of the L8 array that are not to be used in the experiment are blacked out, except for their column headings.

The quadratic loss is:

$$\text{Loss} = (D_{50} - 240)^2$$

The data (D_{50} and Loss) are given in Table 15.3. They are arrayed for each run of the Inner Array, by Outer Array run.

In order to apply the Taguchi decision process, compute the average loss for each level of each control parameter. The levels that yield the lower of the two average losses are the Taguchi choice. Table 15.4 shows the average loss values, the square root of the mean loss as a percentage of target, the ratios of the loss at level 1 to the loss at level 2, and the critical values from the (approximate) doubly non-central F distribution. Note that the parameter values used to compute the critical values were: $\widehat{\mu} = 261.9$ and $\widehat{\sigma} = 50.27$, which are the estimates obtained over all 8 * 4 = 32 observations (i.e., the null hypothesis being that none of the control parameters actually affected the D_{50}).

Table 15.1 L8 inner array—dissolution experiment

T (°C)	C (min)	X3	P (N)	X5	X6	X7
38	30		45			
38	30		55			
38	45		45			
38	45		55			
42	30		45			
42	30		55			
42	45		45			
42	45		55			

Table 15.2 L4 outer array—dissolution experiment

pH	V (mL)
1.5	175
1.5	300
3.5	175
3.5	300

Table 15.3 The dissolution experiment—data

Inner array run	T	C	P	pH	V	*D* 50	Loss
1	38	30	45	1.5	175	211	840.20
2	38	30	55	1.5	175	312	5160.33
3	38	45	45	1.5	175	210	915.23
4	38	45	55	1.5	175	310	4850.13
5	42	30	45	1.5	175	207	1065.15
6	42	30	55	1.5	175	311	5035.05
7	42	45	45	1.5	175	212	789.66
8	42	45	55	1.5	175	311	5065.48
1	38	30	45	1.5	300	205	1232.08
2	38	30	55	1.5	300	309	4753.45
3	38	45	45	1.5	300	211	825.11
4	38	45	55	1.5	300	310	4907.88
5	42	30	45	1.5	300	231	89.82
6	42	30	55	1.5	300	316	5834.86
7	42	45	45	1.5	300	216	575.42
8	42	45	55	1.5	300	316	5839.29
1	38	30	45	3.5	175	208	1026.59
2	38	30	55	3.5	175	308	4576.54
3	38	45	45	3.5	175	219	450.90
4	38	45	55	3.5	175	312	5227.32
5	42	30	45	3.5	175	213	724.06
6	42	30	55	3.5	175	313	5338.02
7	42	45	45	3.5	175	209	967.91
8	42	45	55	3.5	175	318	6116.37
1	38	30	45	3.5	300	209	939.10
2	38	30	55	3.5	300	309	4717.00
3	38	45	45	3.5	300	215	623.48
4	38	45	55	3.5	300	306	4325.93
5	42	30	45	3.5	300	212	760.75
6	42	30	55	3.5	300	305	4270.60
7	42	45	45	3.5	300	213	706.23
8	42	45	55	3.5	300	311	5075.41

Table 15.4 Dissolution experiment—mean loss table

Mean loss				Root mean loss/target		
Level	T	C	P	T	C	P
1	***2835.70***	***2897.73***	***783.23***	***22.2*** %	***22.4*** %	***11.7*** %
2	3015.88	2953.86	5068.35	22.9 %	22.6 %	29.7 %
*L*1/*L*2	0.9403	0.9810	**0.1545**			
F"(0.025)	0.3648	0.3648	0.3648			
F"(0.975)	2.7508	2.7508	2.7508			

The only ratio of $L1/L2$ that was significant at the $\alpha = 0.05$ level was for the control parameter P, the compacting force.

The EAS concludes that setting the compacting force to its low level of 45 N will on the average get the value of D_{50} closest to the target of 240 min with the least amount of variation in the face of variable gastric fluid pH and volume. The values of processing temperature (T) and time (C) may be set to whatever are the most economical values. Using the classical Taguchi approach, the EAS would have chosen the low levels of all three control parameters, since the low levels yielded the minimum mean loss. It is possible that the low levels of T and C are in fact the most economical, but it would not have necessarily been true.

Table 15.5 Mean D_{50}—noise parameter array

pH	V	N	Mean D50
1.5	175	8	260.5
1.5	**300**	**8**	**264.3**
3.5	175	8	262.5
3.5	**300**	**8**	**260.2**

Table 15.6 Mean loss—control parameter array with two noise conditions

V	T	C	P	Mean loss
300	38	30	45	1085.59
300	38	30	55	4735.222
300	38	45	45	724.292
300	38	45	55	4616.903
300	**42**	**30**	**45**	**425.287**
300	42	30	55	5052.729
300	42	45	45	640.826
300	42	45	55	5457.348

Table 15.7 Mean Loss—decision table using only two noise conditions

	Mean loss			Root mean loss/target		
Level	T	C	P	T	C	P
1	***2790.50***	***2824.71***	***719.00***	***22.0*** %	***22.1*** %	***11.2*** %
2	2894.05	2859.84	4965.55	22.4 %	22.3 %	29.4 %
L1/*L2*	0.9642	0.9877	**0.1448**			
F"(0.025)	0.2281	0.2281	0.2281			
F"(0.975)	4.4148	4.4148	4.4148			

Fowlkes and Creveling (1995) suggested an alternative to a factorial outer array design. They suggest using a single combination of control parameter values, and perform an array experiment with only noise parameters. Then find the combination of noise parameters that yields the highest average value of the response (not the loss, or signal-to-noise, but the raw response itself) and the combination of values that yields the lowest average response. Then use only those two sets of noise parameter conditions when running the full inner array. The notion is that:

1. The two conditions chosen would yield the highest variability, so that choosing the control parameter conditions that minimizes the loss in face of the noise conditions would in fact be the most robust design choice;
2. Noise parameters and Control parameters do not interact.

Using the tablet dissolution example, the average value of the response, D50, for each of the noise parameter conditions, is shown in Table 15.5. The conditions that yield the highest and lowest mean D_{50} are $pH = 3.5$, $V = 300$ (lowest mean $D_{50} = 260.2$ min) and $pH = 1.5$, $V = 300$ (highest mean $D_{50} = 264.3$ min). Had only data from those two noise parameter conditions been used, the average loss for the control parameter array would have been as shown in Table 15.6. Finally, the mean loss table for minimizing loss using only the two "extreme" noise conditions is shown in Table 15.7.

This table would lead us to the same conclusion (that $P = 45$ N is optimal, and both parameters T and C could be set to the most economical level) as did Table 15.4 with four (2^2) noise conditions.

Generating Two-Level Orthogonal Arrays

As was mentioned early, all two-level fractional factorials are orthogonal arrays. The methods for generating 2^{k-p} fractional designs have already been discussed. The two-level arrays are generally generated as 2^{k-p} fractions. First, the number of runs, N (a power of 2) is selected. The fractional factorial associated with N is a 2^{k-p} design, with $k = N - 1$. Thus, $2^{k-p} = 2^{N-1-p} = N$, so that $p = N - \log_2 N - 1$. The maximum number of factors that can be used with an LN array is $k = N - 1$.

Consider the L8 array shown in Fig. 15.2. In this case, $N = 8$, so $k = 8 - 1 = 7$ and $p = 8 - \log_2 8 - 1 = 4$. Thus, the L8 array is a 2^{7-4} factional factorial. Table 15.8 shows the values of N, k, and p for N = 4, 8, 16, and 32.

The Taguchi orthogonal arrays are not the only orthogonal arrays that can be generated for N − 1 factors with N runs. The method for generating fractional designs described in Chap. 6 can be used to generate the Taguchi arrays, but the particular design (array) that is generated depends on the choice of generators. Statistically, the 2^{k-p} fractional design chosen does not affect the outcome or method of the analyses.

Generating Three-Level Orthogonal Arrays

Generating three-level arrays is slightly more complex than it is for two-level arrays. The three-level arrays will have numbers of runs equal to powers of 3 (L9 and L27, for example) and the mixed-level arrays will have both 2 and 3 as factors (L12, L18, L36).

The number of runs, N, in a three-level array is equal to 3^{k-p}, where $k = (N - 1)/2$ is the maximum number of factors that can be used with the array. The design parameters are derived in a very similar fashion as those for the two-level arrays. For three-level arrays, $3^{k-p} = 3^{N-1-p} = N$, so that $p = N - \log_3 N - 1$. Table 15.9 shows the parameters N, p, and k for L9, L27, and L81 arrays.

In general, three-level fractional factorial experiments are not the most efficient designs for incorporating a third level. The second-order designs described in Chap. 7 are more efficient. Generating 3^{k-p} fractional designs requires a different kind of coding, where levels are designated as 0, 1, or 2, instead of −1, 0, and +1. Furthermore, interaction effects are not only a function of the

Table 15.8 Equivalence of some orthogonal arrays and two-level fractional factorials

	L4	L8	L16	L32
N	4	8	16	32
$\boldsymbol{k = N - 1}$	3	7	15	31
log2(N)	2	3	4	5
$\boldsymbol{p}$	1	4	11	26
$\boldsymbol{2^{k-p}}$	4	8	16	32

Note: LN array is the same as a 2^{k-p} fractional factorial

Table 15.9 Equivalence of some orthogonal arrays and three-level fractional factorials

	L9	L27	L81
N	9	27	81
$\boldsymbol{k = (N - 1)/2}$	4	13	40
log3(N)	2	3	4
$\boldsymbol{p}$	6	23	76
$\boldsymbol{3^{k-p}}$	9	27	81

Note: LN array is the same as a 3^{k-p} fractional factorial

number of factors interacting, but also the order of the term (either 1 or 2). Montgomery (2001) gives an excellent explanation of how interaction terms are formulated in three-level experiments, and a method for generating 3^{k-p} designs. Hinkelmann and Kempthorne (2005) give a more abstract discussion of the algebraic and combinatorial properties of the three-level fractions and orthogonal arrays. Kacker et al. (1991) give a very thorough, general, and mathematically detailed description of how orthogonal arrays can be generated.

We will use the L9 array to demonstrate the method for generating 3^{k-p} fractional designs, but we will not attempt to generalize the process here.

For an L9 array, consider a 3^{4-2} fractional design. Each term in a four-factor model can be thought of as a product of each factor, with an exponent of 0, 1, or 2 on each factor. So, for example, if X1 is the first factor, then the term:

$$\mathrm{X1}^1\mathrm{X2}^0\mathrm{X3}^0\mathrm{X4}^2 = \mathrm{X1X4}^2$$

Let X1, *X*2, X3, X4 represent the columns of the experimental array, where each row is some combination of 0 s, 1 s, or 2 s.

In order to identify a particular fraction having $3^{4-2} = 3^2 = 9$ runs, consider two unique terms of the form:

$$\mathrm{X1}^{\alpha 1}\mathrm{X2}^{\alpha 2}\mathrm{X3}^{\alpha 3}\mathrm{X4}^{\alpha 4}$$

where the exponents, αi, take the values 0, 1, or 2.

If each run in the full 3^4 experiment is thought of as a column vector, $\boldsymbol{x}$, whose elements are a combination of 0 s, 1, and 2 s, and $\boldsymbol{\alpha}_j$ is the vector:

$$\boldsymbol{\alpha}_j = \begin{bmatrix} \alpha_j 1 \\ \vdots \\ \alpha_j 4 \end{bmatrix}, \; j = 1, 2$$

Then the product $\boldsymbol{\alpha}_j^{'}\boldsymbol{x} = 0,\ 1,$ or 2 (mod 3). The designation "mod 3" means that the product is reduced to its remainder after being divided by 3. To select runs for a 3^{-2} fraction, chose a set of values δ_1 and δ_2, where each of the δ_i equals either 0, 1, or 2. Then for each run in the full 3^4 experiment, compute $\boldsymbol{\alpha}_j^{'}\boldsymbol{x}$ and select those runs where

$$\boldsymbol{\alpha}_j^{'}\boldsymbol{x} = \delta_j \, mod(3), \quad j = 1,\ 2$$

The set of runs where $\delta_1 = \delta_2 = 0$ is called the principal fraction. The equations

$$\boldsymbol{\alpha}_j^{'}\boldsymbol{x} = \delta_j \, mod(3), \quad j = 1,\ 2$$

Are called the Defining Contrasts (Montgomery 2001). As an example consider the set of $\alpha_j(i)$ and δ_j coefficients given in Table 15.10.

Of the $3^4 = 81$ possible runs in a four-factor, three-level experiment, there are only nine satisfying the defining contrasts in Table 15.10. They constitute an L9 array, and are given in Table 15.11.

As in the case of the two-level arrays, the aliasing and resolution of the L9 array depend on the choices of defining contrasts and number of factors ($\leq$4) that are used with this array.

Table 15.10 Coefficients for the defining contrasts of a 3^{4-2} experiment

j	α 1	α 2	α 3	α 4	δ_j
1	1	1	2	0	0
2	1	0	1	1	0

Table 15.11 An L9 array

Pattern	X1	*X2*	X3	X4
1111	0	0	0	0
1223	0	1	1	2
1332	0	2	2	1
2122	1	0	1	1
2231	1	1	2	0
2313	1	2	0	2
3133	2	0	2	2
3212	2	1	0	1
3321	2	2	1	0

```
options nonumber nodate linesize=80 pagesize=54;
libname fact 'H:\Personal Data\Experimentation for Design &
Validation\Data & Analyses';
/* use lib fact to create an output dataset
   containing the design
*/


proc factex;

  factors H1 H2 H3 H4/NLEV=3;
  size fraction=9;
  model res=maximum/;
  examine aliasing;
  output out=frac;
  title1 'Table 15.9 - 3**(4-2) Res III Design';
  run;

  data fact.f20150331_resIII_3_4_design;

    set frac;

      run;
```

Fig. 15.5 SAS PROC FACTEX code for generating an L9 array

There are options for generating three-level fractional designs in JMP, Minitab, and SAS. The SAS PROC FACTEX code for generating an L9 array is given in Fig. 15.5

Table 15.12 shows the array that was generated. SAS uses Helmert coding, so that

$0 \rightarrow -1$

$1 \rightarrow 0$

$2 \rightarrow +1$

Mixed-Level Arrays

Taguchi generated arrays that can accommodate both two-level and three-level factors simultaneously. The numbers of runs in these arrays are multiples of 6 (2 * 3). The most common would be the L12 and L18 arrays. An example of an L12 array, generated with JMP, is given in Table 15.13.

Table 15.12 L9 array generated by SAS PROC FACTEX

Table 15.9 – 3**(4-2) Res III design
The FACTEX procedure
Aliasing structure
H1 = (2*H2) + (2*H3) = H2 + H4 = H3 + (2*H4)
(2*H1) = H2 + H3 = (2*H2) + (2*H4) = (2*H3) + H4
H2 = (2*H1) + (2*H3) = H1 + (2*H4) = H3 + H4
(2*H2) = H1 + H3 = (2*H1) + H4 = (2*H3) + (2*H4)
H3 = (2*H1) + (2*H2) = H1 + H4 = H2 + (2*H4)
(2*H3) = H1 + H2 = (2*H1) + (2*H4) = (2*H2) + H4
H4 = H1 + (2*H2) = (2*H1) + H3 = H2 + (2*H3)
(2*H4) = (2*H1) + H2 = H1 + (2*H3) = (2*H2) + H3

H1	**H2**	**H3**	**H4**
−1	−1	−1	−1
−1	0	1	1
−1	1	0	0
0	−1	1	0
0	0	0	−1
0	1	−1	1
1	−1	0	1
1	0	−1	0
1	1	1	−1

Table 15.13 An L12 array

X1	*X*2	X3	Pattern
1	1	1	−−−
1	1	2	−−0
1	1	3	−−+
1	2	1	− + −
1	2	2	− + 0
1	2	3	−++
2	1	1	+−−
2	1	2	+ − 0
2	1	3	+ − +
2	2	1	++−
2	2	2	0
2	2	3	+++

This L12 array is an L4 array in two 2-level factors, expanded 3 times, once for each level of an additional three-level factor. The L18 designs and their generation are more complex, and will not be discussed. However, it is fairly simple to generate such arrays with the aforementioned software. Table 15.14 shows an L18 array, generated using the Minitab DOE function, with one 2-level factor and three 3-level factors.

Tolerance Design

Once the Parameter Design step is performed, and optimal values for the control parameters are selected, an analysis called Tolerance Design may be performed to help reduce variation in the response. Barker (1990) provides an example of a Tolerance Design analysis following a Parameter Design experiment for selecting optimal control parameter values in an electrical circuit. The Tolerance Design analysis is used to find which control parameters contribute the most to total

Table 15.14 An L18 array

X1	*X2*	X3	X4
1	1	1	1
1	1	2	2
1	1	3	3
1	2	1	1
1	2	2	2
1	2	3	3
1	3	1	2
1	3	2	3
1	3	3	1
2	1	1	3
2	1	2	1
2	1	3	2
2	2	1	2
2	2	2	3
2	2	3	1
2	3	1	3
2	3	2	1
2	3	3	2

variability in the response (impedance). The total variability in the response is related by an equation to the contribution to variability from each control parameter. The equation is used to determine the degree to which each control parameter's contribution to variance should be reduced in order to achieve a desired level of overall variability. To create the equation, Barker first performs what is called a variance components analysis (Searle et al. 2006). The idea is to take the total variance in a set of data and ascribe to each term in a model a proportion of that variance. In order to accomplish this, each term contributing to variance would be identified as a "random effect" in many software packages. Figure 15.6 shows a screen from JMP "Fit Model" function for the tablet manufacturing problem. The control parameters are identified as random effects via the "Attributes" feature. Note that the method for estimating the variance components is set to REML (Restricted Maximum Likelihood). This is the preferred method, as it can guarantee that all the components of variance will be non-negative (a negative variance is an algebraic impossibility, so that any estimation method should reflect this).

Figure 15.7 shows the JMP output for this model.

The column labeled "Pct of Total" gives the contribution of each term to the total variance of the response, D_{50}. The total variance was approximately 119.37, or a total standard deviation of about 10.93 min. If the response was centered at the target of 240 min, then about 95 % of the values would be expected to be somewhere between 240 ± 1.96*10.93, or from 218.58 to 261.42 min. If the EAS wants to reduce the 95 % "range" of D_{50} values, she might want to know how much each control parameter contributes to the total variance, and therefore know which parameters provide the greatest potential for affecting a reduction in the variability. The equation Barker uses would be, for our example:

$$\begin{aligned}\text{Total Variance} &= \text{V}(Total) = 119.37 \\ &= 0.00442*\text{V}(T) + 0.0000*\text{V}(C) + 0.8126*\text{V}(P) + 0.1830*\text{V}(\text{residual})\end{aligned}$$

The variance due to residual error (pure noise) is V(residual) = 21.84. Suppose that a reduction in the 95 % range of D_{50} is desired to be 240 ± 15 min, or (225,255). In other words, the EAS would want $240 - 1.96\sqrt{\text{V}(Total)} = 225$, so V(*Total*) would have to be $V(Total) = \left(\frac{240-225}{1.96}\right)^2 \approx 58.57$. Thus, the reduction in Total Variance requires is 58.57/119.37 ≈ 0.4907. If T^* and P^* represent the required

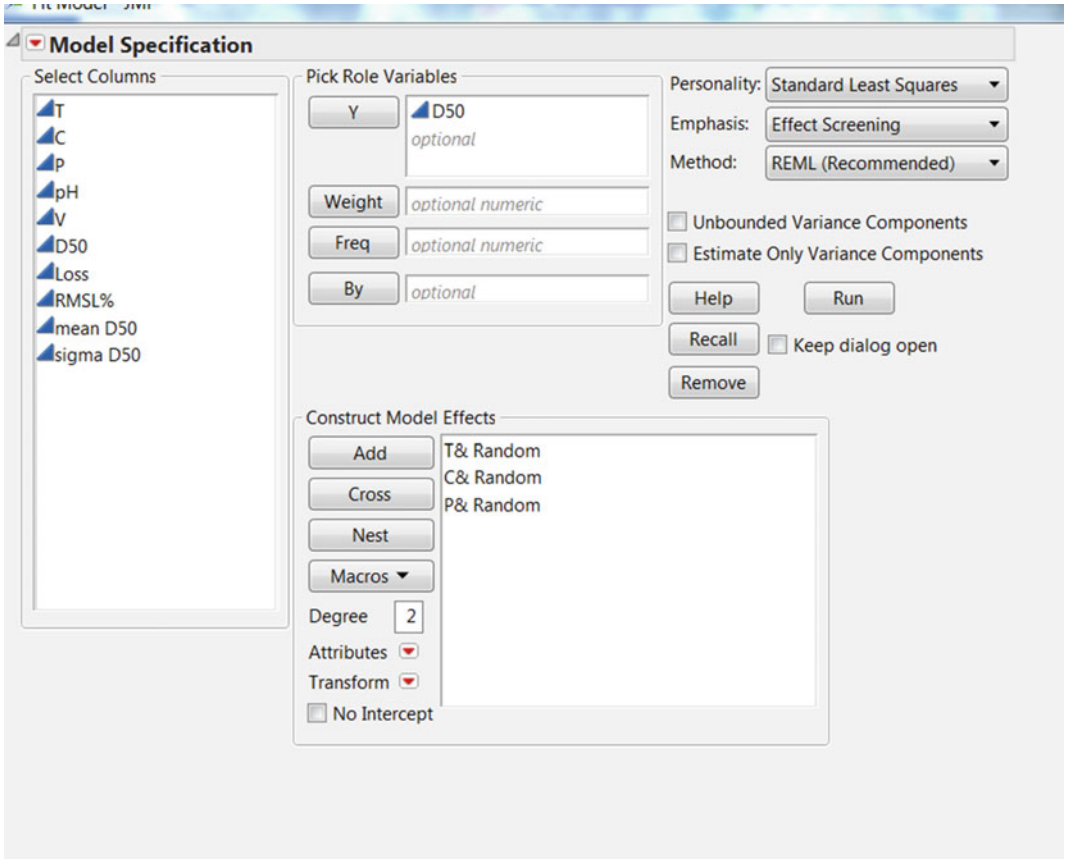


Fig. 15.6 JMP fit model specification for variance components analysis

Response D50

Summary of Fit

RSquare	0.991846
RSquare Adj	0.991846
Root Mean Square Error	4.673684
Mean of Response	261.8544
Observations (or Sum Wgts	32

REML Variance Component Estimates

Random Effect	Var Ratio	Var Component	Std Error	95% Lower	95% Upper	Pct of Total
T	0.0241602	0.5277392	0.988689	0.0814206	91375.121	0.442
C	0	0	0	0	0	0.000
P	4.440541	96.996156	137.21189	19.302047	99138.127	81.259
Residual		21.843319	5.7363367	13.854431	39.474882	18.299
Total		119.36721	137.33314	28.622705	13236.003	100.000

-2 LogLikelihood = 196.62550104

Note: Total is the sum of the positive variance components.

Total including negative estimates = 119.36721

Fig. 15.7 JMP variance components analysis output

fraction of their current variance, and R^* represents residual variance, then the Total Variance Equation could be written as:

$$0.4907 = 0.00442*T^* + 0.8126*P^* + 0.1830*R^*$$

Subtracting out the contribution of residual variance (as a fraction of the Total Variance) gives:

$$0.4907 - 0.1830*21.84/119.37 \approx 0.4572 = 0.00442*T^* + 0.8126*P^*$$

The next step is to find a reduced fraction of variance for either T* or P* to satisfy the equation. Since this is a single equation in two unknowns (fortunately, C contributed nothing to the Total Variance), some trial-and-error is required. Presume for a moment that the fraction of total variance due to T does not change, i.e., $T^* = 1.0000$. Then solve the equation for P^*:

$$P^* = (0.4572 - 0.00442(1))/0.8126 \approx 0.5572$$

That means in order to reach the goal of a 95 % range for D_{50} to be (225, 255), we need to reduce the variance component for P from 96.9961 to 0.5572*96.9961 ≈ 54.0462. Of course, further experimentation with the factor P may be required in order to achieve the desired goal in reduction to variance contribution.

Summary

Taguchi Methods® were developed as a means of using factorial experiments to improve process or product design performance. The designs are sometimes not as efficient as fractional factorials that include both control and noise parameters in a single "array". In particular, the use of three-level Taguchi designs may be much less efficient than second-order response surface designs such as Box–Behnken or Central Composite. While Taguchi used the "signal-to-noise" transformation of the quadratic loss function, it may be easier to use the loss function directly, in light of its more tractable sampling properties. Taguchi's implied presumption is that no control parameter would be included in an experiment if the parameter was not already known to significantly affect the response variable. Nevertheless, Dr. Taguchi provided a relatively easy method to assess alternative product or process designs, and he introduced the notion of a loss function as a design tool.

One element of the discussion about robust design that is glaringly missing is the use of approximating polynomials. Although the traditional Taguchi approach does not include fitting models, least squares can of course be used with the orthogonal array designs. Approximating polynomials might be used to fit the raw response data, or the loss function. In the case of loss, it might be helpful if the logarithm of the absolute value of loss be used as the response, in order to improve the fit and thus the predictive capability of the polynomial.

Key Points

- Dr. Taguchi considered that the appropriate response variable was the degree of loss due to missing a target value.
- He used the quadratic loss function, and transformed it into what he called signal-to-noise ratios (SNRs), so that the objective of design was to maximize SNR rather than minimize loss.
- The loss function has a fairly tractable set of sampling properties, so that it is possible to create confidence intervals for it and compare average loss between two designs.
- Dr. Taguchi recognized that product and process design were multifactorial in nature, and he employed factorial experiments to find optimal values of all the factors or features of the product design.

- Dr. Taguchi also recognized that factors can be classified into two main categories, namely control and noise "parameters". The control parameters are the factors to be optimized in face of the variation possible among the noise parameters.
- The control and noise parameters are varied in two independent experimental arrays. The control array is called the "inner" array, and the noise array is called the "outer" array. The original idea was to run a complete outer array for every run of the inner array.
- A potentially economical method for representing noise without needing to completely replicate the entire noise array for every run of the control parameter array is to run the noise array for one set of control parameter conditions or states, and then find the two noise conditions that provide the minimum and maximum average value of the raw response variable, and using only those two noise conditions in conjunction with the control parameter array.
- The experimental process is performed in two stages, parameter design and tolerance design. Parameter design is for optimization of the control parameter values. Tolerance design analyses helps determine the degree to which contributions to overall variance from each control parameter should be reduced in order to achieve an overall reduction in response variability.

Exercises and Questions

1. An EAS must design a better (humane) mousetrap. There are three control parameters:

   (a) Box height
   (b) Box length
   (c) Door spring strength (controls the speed of closure)

   There are two noise parameters:
   N1) mouse length
   N2) mouse speed
   Assume a mouse may be simulated with a mechanical device having a particular length and speed. What arrays would you suggest?
2. Using the data in Table 15.3, compute a two-sided 95 % confidence interval for the root mean square percent loss when $P = 45$ N (optimal conditions).

References

Adler, J. (2010). *R in a nutshell*. Sebastopol: O'Reilly Media.

Allen, T. T. (2006). *Introduction to engineering statistics and six sigma*. London: Springer.

Armitage, P. (1971). *Statistical methods in medical research*. Oxford: Blackwell Scientific Publications.

Barker, T. B. (1990). *Engineering quality by design*. New York/Milwaukee: Marcel Dekker/Quality Press.

Bickel, P. J., & Doksum, K. A. (2007). *Mathematical statistics: Basic ideas and selected topics* (2nd ed., Vol. 1). Upper Saddle River, NJ: Pearson Prentice Hall.

Box, G. E. P., Hunter, W. G., & Hunter, J. S. (1978). *Statistics for experimenters*. New York: Wiley.

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. New York: Wiley.

Chow, S. C. (2007). *Statistical design and analysis of stability studies*. Boca Raton, FL: Chapman & Hall/CRC Press.

Cochran, W. G., & Cox, D. R. (1992). *Experimental designs* (2nd ed.). New York: Wiley.

Conover, W. J. (1999). *Practical nonparametric statistics* (3rd ed.). New York: Wiley.

Cryer, J. D. (1986). *Time series analysis*. Boston: Duxbury Press.

Desu, M. M., & Raghavarao, D. (1990). *Sample size methodology*. San Diego, CA: Academic.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). New York: Wiley.

Dunn, P. F. (2010). *Measurement and data analysis for engineering and science* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC Press.

Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans*. Philadelphia: SIAM.

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3rd ed.). New York: Wiley.

Fowlkes, W. Y., & Creveling, C. M. (1995). *Engineering methods for robust product design*. Reading, MA: Addison-Wesley.

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

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1997). *Bayesian data analysis*. Boca Raton, FL: Chapman & Hall/CRC Press.

Grant, E. L., & Leavenworth, R. S. (1980). *Statistical quality control* (5th ed.). New York: McGraw-Hill.

Hinkelmann, K., & Kempthorne, O. (2005). *Design and analysis of experiments: Vol. 2. Advanced experimental design*. Hoboken, NJ: Wiley.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York: Wiley.

Incropera, F. P., & De Witt, D. P. (1990). *Introduction to heat transfer* (2nd ed.). New York: Wiley.

Johnson, N. L. (1949). Systems of frequency curves generated by translation. *Biometrika, 36*, 149–176.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions* (2nd ed., Vol. 2). New York: Wiley.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H., & Lee, T. C. (1985). *The theory and practice of econometrics* (2nd ed.). New York: Wiley.

Kacker, R. N., Lagergren, E. S., & Filliben, J. J. (1991). Taguchi's orthogonal arrays are classical designs of experiments. *Journal of Research of the National Institute of Standards and Technology, 96*(5), 577–591.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53*, 457–481.

Law, A. M., & Kelton, W. D. (1982). *Simulation modeling and analysis*. New York: McGraw-Hill.

Lee, E. T. (1992). *Statistical methods for survival data analysis* (2nd ed.). New York: Wiley.

Mann, N. R., Schafer, R. E., & Singpurwalla, N. D. (1974). *Methods for statistical analysis of reliability and life data*. New York: John Wiley and Sons.

S.A. Pardo, *Empirical Modeling and Data Analysis for Engineers and Applied Scientists*,
DOI 10.1007/978-3-319-32768-6

Meyer, P. L. (1970). *Introduction to probability and statistical applications* (2nd ed.). Boston: Addison-Wesley.
Montgomery, D. C. (2001). *Design and analysis of experiments*. New York: Wiley.
Nash, R. A., & Wachter, A. H. (2007). *Pharmaceutical process validation* (3rd ed.). New York: Informa Healthcare.
Otto, K., & Wood, K. (2001). *Product design: Techniques in reverse engineering and New product development*. Upper Saddle River, NJ: Prentice-Hall.
Pardo, S. A. (2009). A differential equation model for dissolution profiles, and its uses in designing dosage forms. *Statistics in Biopharmaceutical Research, 1*(2), 194–200.
Pardo, S. A. (2014). *Equivalence and noninferiority tests for quality, manufacturing, and test engineers*. Boca Raton, FL: Chapman & Hall/CRC Press.
Rutherford, E. (1900). A radioactive substance emitted from thorium compounds. *Philosophical Magazine Series 5, XLIX*, 1–14.
Seber, G. A. F., & Wild, C. J. (1989). *Nonlinear regression*. New York: Wiley.
Searle, S. R., Casella, G., & Mcculloch, C. E. (2006). *Variance components*. New York: John Wiley and Sons.
Smith, W. F. (2005). *Experimental design for formulation*. Philadelphia, PA: SIAM/ASA.
Springer, M. D. (1979). *The algebra of random variables*. New York: Wiley.
Taguchi, G., & Wu, Y. (1980). *Introduction to off-line quality control*. Nagoya: Central Japan Quality Control Association.
White, G. W. (2012). The design of natural gas pipelines. *Pipeline and Gas Journal, 239*(9) (Sept. 2012), http://pgjonline.com.
Whitten, K. W., Davis, R. E., Peck, M. L., & Stanley, G. G. (2004). *General chemistry* (7th ed.). Belmont, CA: Brooks/Cole.

Index

S.A. Pardo, *Empirical Modeling and Data Analysis for Engineers and Applied Scientists*,
DOI 10.1007/978-3-319-32768-6