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.