

Use R!

Series Editors:

Robert Gentleman • Kurt Hornik • Giovanni Parmigiani

Series Editors:

Robert Gentleman
Program in Computational Biology
Division of Public Health Sciences
Fred Hutchinson Cancer Research Center
1100 Fairview Ave. N, M2-B876
Seattle, Washington 98109-1024
USA

Kurt Hornik
Department für Statistik und Mathematik
Wirtschaftsuniversität Wien Augasse 2-6
A-1090 Wien
Austria

Giovanni Parmigiani
The Sidney Kimmel Comprehensive Cancer
Center at Johns Hopkins University
550 North Broadway
Baltimore, MD 21205-2011
USA

For further volumes:

<http://www.springer.com/series/6991>

Use R!

Albert: Bayesian Computation with R

Bivand/Pebesma/Gómez-Rubio: Applied Spatial Data Analysis with R

Cook/Swayne: Interactive and Dynamic Graphics for Data Analysis:
With R and GGobi

Hahne/Huber/Gentleman/Falcon: Bioconductor Case Studies

Paradis: Analysis of Phylogenetics and Evolution with R

Pfaff: Analysis of Integrated and Cointegrated Time Series with R

Sarkar: Lattice: Multivariate Data Visualization with R

Spector: Data Manipulation with R

Emilio L. Cano • Javier M. Moguerza
Andrés Redchuk

Six Sigma with R

Statistical Engineering for Process
Improvement

 Springer

Emilio L. Cano
Department of Statistics and Operations
Research
Rey Juan Carlos University
Madrid, Spain

Javier M. Moguerza
Department of Statistics and Operations
Research
Rey Juan Carlos University
Madrid, Spain

Andrés Redchuk
Department of Statistics and Operations
Research
Rey Juan Carlos University
Madrid, Spain

ISBN 978-1-4614-3651-5 ISBN 978-1-4614-3652-2 (eBook)
DOI 10.1007/978-1-4614-3652-2
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012939055

© Springer Science+Business Media New York 2012

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To our families

Foreword

Many books on Six Sigma have appeared in recent decades. Most of them follow a standard structure: a description of Six Sigma plus some examples of its successful implementation. And many of them have the same motivation: through Six Sigma you will achieve huge cost reductions in your organization, quantified as being between 10 and 30%. It is an appealing claim, especially in the current tough economic climate, but it is an empirical one; as such, it is hard to prove and, of course, a gateway to criticism. You will not find that kind of argument in this book. Its motivation is purely scientific: Six Sigma is “a quality paradigm that translates the involved scientific methodology into a simple way to apply the scientific method.” And it really does. When selling a concept within an organization, make sure that the concept is understood. For decades, scientists have been trying to sell great ideas to industry with less than stellar success. In my opinion, this is because, too frequently, we were speaking two different languages. This is a hard obstacle to overcome. Six Sigma has emerged as one of the solutions to this lack of understanding.

This solution did not arise from the scientific camp: it came directly from Motorola and General Electric, two leading industrial organizations. Yet there is a lesson in this for scientists: If you want to be understood by industry, use the language of Six Sigma (or a similar one). And, of course, take Six Sigma as an example of how scientists should work to transfer knowledge and technology to industry.

This book deals with the use of R within Six Sigma. R is an open platform widely used in the academic and scientific communities. Its success is based on two main features: (1) R is robust, rigorous, and efficient and (2) R is free. The latter is a real example of how to reduce costs: if your company uses expensive software tools, then this book is an opportunity to reduce them. It’s that simple.

But the book is addressed not just to Six Sigma practitioners. If you do not know what Six Sigma is, then this could be the book for you. The Six Sigma philosophy is introduced from the beginning, making the book a self-contained work. In fact, I strongly recommend the first chapter as a simple and complete description of Six Sigma. All the facts are there.

The authors have combined their scientific, technological, industrial, and pedagogical skills in this one volume. I know them well.

Prof. Javier Moguerza is an accomplished multidisciplinary scientist. He holds active membership in the Global Young Academy. As part of his scientific backpack, he is a specialist in project management. I have seen him work with heterogeneous groups, always successfully. Applying Six Sigma is a natural way of thinking for him. He is “cause-and-effect born” or, if you prefer, “Six Sigma born.”

Dr. Andres Redchuk is a Six Sigma Master Black Belt (read the book’s first chapter to find out what this means). For almost 20 years, he has been working as a Six Sigma practitioner in the private sector, with more than a hundred projects under his belt. He informs the book’s industry point of view.

Emilio Lopez is one of the best Ph.D. students we have seen in the Department of Statistics and Operations Research at Universidad Rey Juan Carlos. He brings broad industry experience and is currently applying Six Sigma as a tool for academic improvement.

The result of these authors’ blending is a perfect conciliation of the scientific and industry points of view. For scientists, the book will describe the needs of the industrial sector. Industrial practitioners, on the other hand, will find an innovative, rigorous, and economical way to apply Six Sigma.

Writing a book about Six Sigma is not an easy task. Writing a useful and original book on Six Sigma is an almost Herculean undertaking. I hope you will benefit from the fruits of the author’s labor.

Royal Academy of Sciences, Spain

David Rios

Preface

Why Six Sigma with R

Six Sigma has grown over the last two decades as a breakthrough Total Quality Management methodology. With Six Sigma, problems are solved and processes are improved, taking as a basis one of the most powerful tools of human development: the scientific method. For the analysis of data, Six Sigma requires the use of statistical software, with R an open source option that fulfills this requirement. R is a software system that includes a programming language widely used in academic and research departments. It is currently becoming a real alternative within corporate environments.

Many publications deal with R or Six Sigma separately, but none addresses the use of statistical techniques with R while at the same time focusing on the Six Sigma methodology. Thus, we were encouraged to write this book and develop an R package as a contribution to the R Project. The aim of this book is to show how R can be used as a software tool in the development of Six Sigma projects.

Who Is This Book For?

This book is not intended as a very advanced or technical manual. It is aimed at addressing the interests of a wide range of readers, providing something interesting to everybody. To achieve this objective, we have tried to include as few equations and formulas as possible. When we find it necessary to use them, we follow them up with simple numerical examples to make them understandable to someone with basic math skills. The examples clarify the tools presented using the Six Sigma language and trying to convey the Six Sigma philosophy.

As far as the software is concerned, we have not used complicated programming structures. Most examples follow the structure `function(arguments) → results`. In this regard, the book is self-contained because it includes all the

necessary background concepts. Nevertheless, we provide plenty of references to both generic and specific R books.

Six Sigma practitioners will find a roadmap at the beginning of the main parts of the book indicating the stage of Six Sigma about to be treated. Moreover, the book contains a chapter with an overview of the R system and a basic reference guide as an appendix. Although you may know about Six Sigma, it is advisable to read the first chapter and check some of the references therein.

Statistical software users and programmers working in organizations using Six Sigma and related methodologies will find in this book a useful alternative. Similarly, analysts and advisers of consulting firms will learn new approaches to their businesses.

Statistics teachers have in a single book the essentials of both disciplines (Six Sigma and R). Thus, the book can be used as a textbook or reference book for intermediate-level courses in engineering statistics, quality control, or related topics.

Finally, business managers who wish to understand and acquire the proper background to encourage their teams to improve their business through Six Sigma can read selected chapters or sections of the book, focusing on the examples.

How to Read This Book

In this book, we present the main tools used in the Six Sigma methodology and explain how to implement them using R. Our intention is to show the relevance of the Six Sigma work flow, the so-called DMAIC cycle (design, measure, analyze, improve, control). This is why, in some parts of the book, concepts that will be defined in subsequent chapters are intuitively used in advance. Detailed description of the concepts can be consulted using the subject index at the end of the book.

The book is organized into seven parts. Part I contains two chapters with the necessary background information for both Six Sigma and R. Parts II–IV constitute the path for our intuitive roadmap. This roadmap goes over the DMAIC cycle throughout its five phases, relating each part to one phase and the main tasks and tools within it. Finally, Part VII comprises a short description of other important tools that the reader should know to have a global view of the advanced possibilities of both Six Sigma and R. An appendix, organized in tables with the main functions used throughout the book, completes the contents.

The chapters have a common structure with an introduction to the given topic, followed by an explanation illustrated with straightforward and reproducible examples. The material used in these examples (data and code) as well as the results (output and graphics) are included sequentially as the concepts are explained. All figures include a brief explanation to enhance the understanding of the interpretation. The last section of each chapter includes a summary and references useful to extend the chapter contents. Finally, some exercises are proposed. The reader is invited to apply the chapter contents on the helicopter case study presented in the first chapter. It is a common example used in quality courses and literature.

A brief guidance is given, leaving freedom to the reader to deal with it in depth at his/her convenience. Finally, we propose some exercises with solutions at the end of the book.

Though the book may be read from beginning to end for a straightforward comprehension of the DMAIC cycle, it can be used also as a reference manual. You can go directly to a specific chapter to learn (or remember) how to perform certain tasks. Thus, you may encounter the same topic explained in more than one way within the book.

We are aware that the book’s contents are only a part of the large number of existing tools, methods, models, and approaches. It was not our intention to write the “Bible of Six Sigma with R” as this would deserve several volumes. The book paves the way to encouraging readers to delve into Six Sigma and R in depth, perhaps igniting an enthusiasm for both topics that is as strong in the readers as it is in the authors. To this end, a number of references are provided in each chapter. With the background acquired in the R system and the references provided (Internet-based articles and books), it will be easy for Six Sigma practitioners to extend the use of R to other tools not treated in the book.

Conventions

We use a homogeneous typeset throughout the book so that elements can be easily identified by the reader. Text in a **sans serif** font relates to software (e.g., R, Minitab). Text in **teletype** font within paragraphs is used for R components (packages, functions, arguments, objects, commands, variables, etc.).

The commands and scripts are also in **teletype** font, preceded by a “>” symbol and enclosed in a gray box with a left darker line:

```
> #This is an input code example
> my.var <- rnorm(10, 2, 0.5)
> summary(my.var)
```

The outputs appear just below the command that produces it, enclosed in a gray box (lighter than the input) and a thin border:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.8231	1.6790	2.1000	1.9640	2.3820	2.5600

The book contains quite a few examples. They start with the string *Example* (*Brief title for the example*). and end with a square (□) below the last line of the example. In the subsequent continuation of the examples within the chapters, the string (*cont.*) is added to the example title.

In the book we will refer to a *Black Belt* as a person in charge of a Six Sigma project. Black Belt and other roles are defined in Chap. 1. Perhaps there will be tasks in the book (especially in the examples) that are not carried out by a Black

Belt; however, we will always use this name for the sake of simplicity. Similarly, what we say about products will very often be suitable for services, and we use in a general manner the term *customer* when referring to customers or clients.

Production

The book was written in `.Rnw` files. The Eclipse IDE (Integrated Development Environment) with the plug-in StatET was used as both editor and interface with R. If you have a different version of R or an updated version of the packages, you might not obtain exactly the same outputs. The session info of the machine where the code has been run is as follows:

- R version 2.14.1 (2011-12-22), i686-pc-linux-gnu
- Locale: LC_CTYPE=en_US.UTF-8, LC_NUMERIC=C, LC_TIME=es_ES.UTF-8, LC_COLLATE=en_US.UTF-8, LC_MONETARY=es_ES.UTF-8, LC_MESSAGES=en_US.UTF-8, LC_PAPER=es_ES.UTF-8, LC_NAME=es_ES.UTF-8, LC_ADDRESS=es_ES.UTF-8, LC_TELEPHONE=es_ES.UTF-8, LC_MEASUREMENT=es_ES.UTF-8, LC_IDENTIFICATION=es_ES.UTF-8
- Base packages: base, datasets, graphics, grDevices, grid, methods, stats, utils
- Other packages: ggplot2 0.8.9, lattice 0.20-0, MASS 7.3-16, nortest 1.0, plyr 1.7, proto 0.3-9.2, qcc 2.2, qualityTools 1.50, reshape 0.8.4, rj 1.0.3-7, SixSigma 0.5.0, xtable 1.6-0
- Loaded via a namespace (and not attached): rj.gd 1.0.3-3, tools 2.14.1

Resources

The code and figures included in the book are available at the book's Web site: <http://www.SixSigmaWithR.com>. The SixSigma package and the data sets can also be downloaded. Links and materials will be updated regularly.

About the Authors

Emilio L. Cano is Adjunct Lecturer in the Department of Mathematics at the University of Castilla-La Mancha and Research Assistant Professor in the Department of Statistics and Operations Research at Universidad Rey Juan Carlos. He has more than 14 years of experience in the private sector as statistician.

Javier M. Moguerza is Associate Professor in Statistics and Operations Research at Universidad Rey Juan Carlos. He publishes mainly in the fields of mathematical programming and machine learning. Currently, he is leading national and international research ICT projects funded by public and private organizations. He has held membership in the Global Young Academy since 2010.

Andres Redchuk is Master Black Belt and Research Assistant Professor in the Department of Statistics and Operations Research at Universidad Rey Juan Carlos. He has worked for more than 18 years as a senior consultant in the private sector.

Madrid, Spain

Emilio L. Cano
Javier M. Moguerza
Andres Redchuk

Acknowledgements

We wish to thank David Rios Insua for his kind foreword and for the time he devoted to reading the manuscript. Mariano Prieto also provided valuable comments on our work, and his publications and courses applying Six Sigma have been an inspiration for us. We appreciate the gentle review of Virgilio Gomez-Rubio as an experienced R author and contributor. Our colleague Javier Cano has generously supported us with some book production issues. We thank the Springer staff (Kurt Hornik, Mark Strauss, Hannah Bracken), who from the beginning have lent their enthusiastic support to our proposal and have seen the project through to completion. A debt of gratitude must be paid to R contributors, particularly to the R core group (<http://www.r-project.org/contributors.html>), for their huge work in developing and maintaining the R project. Most of the quotes included throughout the book can be found at <http://www.wikiquote.org/>. We also acknowledge projects VRTUOSI (LLP, code 502869-LLP-1-2009-ES-ERASMUS-EVC) and RIESGOS-CM (code S2009/ESP-1685), in which the methodology described in this book has been applied.

Last but not least, we will be eternally grateful to our families for their patience and for forgiving us for the stolen time. Thank you Alicia, Angela, Erika, Lucia, Pablo, Rita, Sonia, Sophie, and Vicky.

Contents

Part I Basics

1	Six Sigma in a Nutshell	3
1.1	Introduction	3
1.2	Brief History	3
1.3	What Is Six Sigma?	4
1.4	DMAIC Cycle	5
1.4.1	Define Phase	6
1.4.2	Measure Phase	6
1.4.3	Analyze Phase	6
1.4.4	Improve Phase	7
1.4.5	Control Phase	7
1.5	Six Sigma Operational Structure	8
1.6	Summary and Further Reading	9
	Case Study	12
2	R from the Beginning	15
2.1	Introduction	15
2.2	First Steps	16
2.2.1	Get and Install	16
2.2.2	Run and Interact	19
2.2.3	Ask for Help	24
2.3	Coping with Data	25
2.3.1	Data Types	25
2.3.2	Creating Data Objects	25
2.3.3	Accessing Data	28
2.3.4	Importing and Exporting Data	30
2.4	Objects and Functions	32
2.4.1	Objects	32
2.4.2	Functions	34

- 2.5 Operators and Functions Commonly Used..... 36
 - 2.5.1 Operators 36
 - 2.5.2 Mathematical Functions 36
 - 2.5.3 Functions for Vectors 37
 - 2.5.4 Loop and Summary Functions 37
- 2.6 Graphics in R 38
 - 2.6.1 Plotting Functions 39
 - 2.6.2 Bivariate Plots 40
 - 2.6.3 Pie Plots 42
- 2.7 Statistics 43
 - 2.7.1 Samples and Combinatorial Computations..... 43
 - 2.7.2 Random Variables 44
- 2.8 Summary and Further Reading 45
- Case Study 46
- Practice 46

Part II R Tools for the Define Phase

- 3 Process Mapping with R 51**
 - 3.1 Introduction 51
 - 3.2 Process Mapping as a Problem-solving Method 51
 - 3.3 Strategies for Process Mapping 52
 - 3.3.1 Two-stage Process Mapping 53
 - 3.3.2 Drilling Down into the Process Steps..... 54
 - 3.4 Step-by-Step Process Mapping 55
 - 3.4.1 Identifying Inputs and Outputs..... 55
 - 3.4.2 Listing the Project Steps 55
 - 3.4.3 Identifying the Outputs of Each Step 56
 - 3.4.4 Identifying the Parameters of Each Step 56
 - 3.4.5 Classifying the Parameters 56
 - 3.5 Drawing a Process Map with the Six Sigma Package 58
 - 3.6 Why Should I Use R for Drawing Diagrams? 58
 - 3.7 Summary and Further Reading 60
 - Case Study 60
 - Practice 61
- 4 Loss Function Analysis with R 63**
 - 4.1 Introduction 63
 - 4.2 Cost of Poor Quality 64
 - 4.3 Modeling the Loss Function 65
 - 4.3.1 Some Notation 65
 - 4.3.2 Taguchi Loss Function 66
 - 4.4 Average Loss Function 69
 - 4.5 Use of Loss Function Within DMAIC Cycle 69
 - 4.6 Loss Function Analysis with Six Sigma Package..... 70

- 4.7 Other Models 71
- 4.8 Summary and Further Reading 73
- Case Study 74
- Practice 75

Part III R Tools for the Measure Phase

- 5 Measurement System Analysis with R** 79
 - 5.1 Introduction 79
 - 5.1.1 Definitions 79
 - 5.2 Data Analysis 80
 - 5.2.1 Data Collection 80
 - 5.2.2 First Approach to Analysis of Variance (ANOVA) 82
 - 5.2.3 Assessing the Measurement System 84
 - 5.3 Using the SixSigma Package 85
 - 5.3.1 Interpreting the Charts 88
 - 5.4 Summary and Further Reading 89
 - Case Study 90
 - Practice 90
- 6 Pareto Analysis with R** 91
 - 6.1 Introduction 91
 - 6.2 Pareto Principle 91
 - 6.2.1 Pareto Principle as a Problem-solving Technique 92
 - 6.3 Pareto Analysis in Six Sigma 93
 - 6.3.1 Identifying Causes 93
 - 6.3.2 Measuring the Effect 95
 - 6.3.3 Building a Pareto Chart 96
 - 6.4 Pareto Charts in R 96
 - 6.5 Other Uses of the Pareto Chart 99
 - 6.6 Summary and Further Reading 100
 - Case Study 100
 - Practice 100
- 7 Process Capability Analysis with R** 101
 - 7.1 Introduction 101
 - 7.2 Specifications 101
 - 7.3 Process Performance 102
 - 7.4 Process vs. Specifications 106
 - 7.5 Capability Indices 108
 - 7.6 Capability Study with SixSigma Package 110
 - 7.7 Summary and Further Reading 111
 - Case Study 112
 - Practice 112

Part IV R Tools for the Analyze Phase

8	Charts with R	115
8.1	Introduction	115
8.1.1	Use of Charts	115
8.1.2	Background Concepts	116
8.2	Bar Chart	119
8.3	Histogram	120
8.4	Scatterplot	124
8.5	Run Chart	126
8.6	Tier Chart	129
8.7	Box–Whisker Chart	130
8.8	Other Charts	132
8.8.1	Group Chart	133
8.8.2	Location Charts	134
8.8.3	Multivariate Chart	134
8.8.4	More About Charts	136
8.9	Summary and Further Reading	138
	Case Study	138
	Practice	139
9	Statistics and Probability with R	141
9.1	Introduction	141
9.1.1	Variables and Observations	142
9.1.2	Summary Tables	143
9.1.3	Population and Sample	144
9.1.4	Special Data Values	145
9.2	Descriptive	146
9.2.1	Measures of Central Tendency	146
9.2.2	Measures of Variability	147
9.3	Probability	150
9.3.1	Random Variables	152
9.3.2	Binomial Distribution	152
9.3.3	Normal Distribution	154
9.3.4	Other Useful Distributions	157
9.4	Summary and Further Reading	164
	Case Study	165
	Practice	165
10	Statistical Inference with R	167
10.1	Introduction	167
10.2	Confidence Intervals	167
10.2.1	Sampling Distribution and Point Estimation	167
10.2.2	Proportion Confidence Interval	171
10.2.3	Mean Confidence Interval	172
10.2.4	Variance Confidence Interval	173

- 10.3 Hypothesis Testing 175
- 10.4 Regression 179
 - 10.4.1 Model Identification 179
 - 10.4.2 Model Fitting 181
 - 10.4.3 Model Validation 184
 - 10.4.4 Other Models 185
- 10.5 Analysis of Variance 187
 - 10.5.1 Model Identification 188
 - 10.5.2 Model Fitting and Validation 189
 - 10.5.3 Additional Models and Related Tools 191
- 10.6 Summary and Further Reading 192
- Case Study 193
- Practice 193

Part V R Tools for the Improve Phase

- 11 Design of Experiments with R** 197
 - 11.1 Introduction 197
 - 11.2 Importance of Experimenting 197
 - 11.2.1 Inconsistent Data 198
 - 11.2.2 Variable Value Range 198
 - 11.2.3 Correlated Variables 198
 - 11.3 Experimentation Strategies 200
 - 11.3.1 Planning Strategies 200
 - 11.3.2 Factor Levels and Replications 201
 - 11.3.3 Progressive Experimentation 202
 - 11.3.4 Model Assumptions 202
 - 11.4 2^k Factorial Designs 203
 - 11.5 Design of Experiments for Process Improvement 212
 - 11.6 Summary and Further Reading 214
 - Case Study 215
 - Practice 215

Part VI R Tools for the Control Phase

- 12 Process Control with R** 219
 - 12.1 Introduction 219
 - 12.2 Mistake-proofing Strategies (Poka-Yoke) 219
 - 12.3 What to Control 220
 - 12.4 Control Chart Basics 221
 - 12.4.1 The Chart 221
 - 12.4.2 Chart Interpretation 221
 - 12.4.3 Sampling Strategy 223
 - 12.5 Plotting Process Control Charts 223
 - 12.5.1 Notation for Computing Control Lines 225

- 12.5.2 Variable Control Charts 226
- 12.5.3 Attribute Control Charts 234
- 12.6 Summary and Further Reading 237
- Case Study 237
- Practice 238

Part VII Further and Beyond

- 13 Other Tools and Methodologies 241**
 - 13.1 Introduction 241
 - 13.2 Failure Mode, Effects, and Criticality Analysis 241
 - 13.3 Design for Six Sigma 242
 - 13.4 Lean 243
 - 13.5 Gantt Chart 243
 - 13.6 Some Advanced R Topics 244
 - 13.6.1 Programming 244
 - 13.6.2 R User Interfaces 246
 - 13.6.3 Reporting 247
 - 13.7 Summary and Further Reading 249
 - Case Study 249
- A R Basic Reference Guide 251**
- References 261**
- Solutions 267**
- R Packages and Functions Used in the Book 275**
- Subject Index 279**

List of Figures

Fig. 1.1	Relationship between DMAIC cycle and scientific method.....	8
Fig. 1.2	Roadmap of DMAIC cycle	8
Fig. 1.3	Strategic Six Sigma roles	10
Fig. 1.4	Operational Six Sigma roles	10
Fig. 1.5	Paper helicopter template	11
Fig. 1.6	Paper helicopter mounting scheme	12
Fig. 2.1	CRAN Web page	17
Fig. 2.2	CRAN Web page. Selection of subdirectory	17
Fig. 2.3	CRAN Web page. Download link for Windows	18
Fig. 2.4	R installation wizard	18
Fig. 2.5	R's graphical user interface	19
Fig. 2.6	R text output.....	20
Fig. 2.7	R graphic output	21
Fig. 2.8	R script editor	22
Fig. 2.9	Plotting curves and math expressions	39
Fig. 2.10	A basic scatterplot.....	40
Fig. 2.11	Improved scatterplot	41
Fig. 2.12	Box plot chart	42
Fig. 2.13	Bar plot	42
Fig. 2.14	Histogram with density line	43
Fig. 2.15	Pie plot.....	43
Fig. 3.1	A problem-solving flow	52
Fig. 3.2	SIPOC flow	53
Fig. 3.3	Top-level process map for pizza example	54
Fig. 3.4	Process map for pizza example	59
Fig. 4.1	Frequency distribution of observed values and requirement limits	64
Fig. 4.2	Cost of poor quality under classical approach	65
Fig. 4.3	Taguchi loss function	67

Fig. 4.4	Taguchi loss function for bolt example.....	68
Fig. 4.5	Loss function analysis for bolt example.....	72
Fig. 4.6	Smaller-the-better loss function	73
Fig. 4.7	Larger-the-better loss function.....	74
Fig. 5.1	Variability decomposition.....	82
Fig. 5.2	Charts for MSA of battery example	87
Fig. 6.1	Distribution of causes in a Pareto analysis	93
Fig. 6.2	Cause-and-effect diagram for construction example	94
Fig. 6.3	Pareto chart with base graphics.....	97
Fig. 6.4	Pareto chart in <code>qcc</code> package.....	98
Fig. 6.5	Pareto chart in <code>qualityTools</code> package.....	99
Fig. 7.1	Specification limits	103
Fig. 7.2	Taguchi's loss function.....	104
Fig. 7.3	Natural tolerance limits or effective limits. Between the mean and three standard deviations lie 99.7% of data.....	109
Fig. 7.4	Capability study output	111
Fig. 8.1	Two-dimensional chart foundation	116
Fig. 8.2	Different visualizations for the same data.....	117
Fig. 8.3	Bar plot of defects by region (printer cartridge example)	119
Fig. 8.4	Multiple bar plot for printer cartridge example	121
Fig. 8.5	Various histogram patterns.....	122
Fig. 8.6	Histograms for printer cartridge example	123
Fig. 8.7	Histogram with density lines (cartridge example)	124
Fig. 8.8	Scatterplot patterns	125
Fig. 8.9	Scatterplot for printer cartridge example.....	126
Fig. 8.10	Patterns in a run chart	127
Fig. 8.11	Run chart for printer cartridge example	128
Fig. 8.12	Tier chart for printer cartridge example	129
Fig. 8.13	Box-whisker charts for printer cartridge example.....	131
Fig. 8.14	Labeling outliers in a box plot.....	132
Fig. 8.15	Group chart for printer cartridge example	134
Fig. 8.16	Multivariate chart for printer cartridge big data frame	137
Fig. 9.1	Long-term and short-term variability	149
Fig. 9.2	Normal distribution	154
Fig. 9.3	Looking for the normal shape	155
Fig. 9.4	Quantile-quantile plot	156
Fig. 9.5	Hypergeometric distribution	157
Fig. 9.6	Geometric distribution	158
Fig. 9.7	Negative binomial distribution	159
Fig. 9.8	Poisson distribution	159
Fig. 9.9	Exponential distribution	160
Fig. 9.10	Lognormal distribution.....	160

Fig. 9.11	Uniform distribution	161
Fig. 9.12	Gamma distribution	161
Fig. 9.13	Beta distribution	162
Fig. 9.14	Weibull distribution	163
Fig. 9.15	Student's t distribution	163
Fig. 9.16	Chi-square (χ^2) distribution	164
Fig. 9.17	F distribution	165
Fig. 10.1	Confidence interval for guitar string example	175
Fig. 10.2	Basic regression scatterplot	181
Fig. 10.3	Sophisticated scatterplot using <code>ggplot2</code> package	182
Fig. 10.4	Graphic output for linear regression model	185
Fig. 10.5	Group box plots for guitar string example	188
Fig. 10.6	Plot of effects in ANOVA for guitar string example	191
Fig. 10.7	Effects of string type	192
Fig. 11.1	Value range risk in observational studies	199
Fig. 11.2	Risks of correlated variables without experimentation	199
Fig. 11.3	Simple effect plot	209
Fig. 11.4	Factor effects in pizza experiment	210
Fig. 11.5	Plot of interactions between factors	211
Fig. 11.6	Model diagnostics for pizza experiment	212
Fig. 12.1	Control charts vs. probability distribution	222
Fig. 12.2	Identifying special causes through individual points	222
Fig. 12.3	Patterns in control charts	224
Fig. 12.4	Decision tree for basic process control charts	225
Fig. 12.5	Individual control chart	228
Fig. 12.6	Moving-range control chart	232
Fig. 12.7	\bar{x} control chart	233
Fig. 12.8	p chart for particle board example	236
Fig. 13.1	Gantt chart	245
Fig. 13.2	PDF file with design matrix	248

List of Tables

Table 2.1	R Console shortcuts	20
Table 2.2	Special data values in R	25
Table 2.3	Main R data types.....	25
Table 2.4	Arguments for <code>matrix</code> function	34
Table 2.5	Data for practice 1, Chap. 2	47
Table 3.1	Arguments of <code>ss.pMap</code> function	58
Table 4.1	Data for bolt example	69
Table 4.2	Arguments for <code>ss.lfa</code> function	70
Table 4.3	Values returned by <code>ss.lfa</code> function.....	70
Table 5.1	Battery voltage measures	81
Table 5.2	Arguments for <code>ss.rr</code> function.....	85
Table 5.3	Practice: Compound in a pastry batch	90
Table 7.1	Arguments for <code>ss.ca.yield</code> function.....	106
Table 7.2	DPMO through sigma scores.....	107
Table 7.3	Arguments for <code>ss.ca.z</code> function.....	107
Table 7.4	Data for winery example	108
Table 7.5	Arguments for <code>ss.ca.study</code> function	110
Table 8.1	Data for printer cartridge example	117
Table 8.2	Defects by region for printer cartridge example.....	120
Table 11.1	Pizza dough recipe scores	205
Table 11.2	Data for pizza dough example.....	205
Table 11.3	Robust experimental data	214
Table 12.1	Data for p chart example	235
Table 13.1	FMECA risk matrix.....	242
Table A.1	Help functions.....	251
Table A.2	Functions for packages and workspace	251

Table A.3	Constants and special values	252
Table A.4	Operators for objects	252
Table A.5	Functions for data types	252
Table A.6	Functions for importing and exporting data	253
Table A.7	Functions for managing data	253
Table A.8	String functions	253
Table A.9	Functions for exploring data	254
Table A.10	Operators	254
Table A.11	Set operations	254
Table A.12	Mathematical functions	254
Table A.13	Vector functions	255
Table A.14	Summarize functions	255
Table A.15	Probability distributions	256
Table A.16	Sampling and combinatorial	257
Table A.17	Graphic functions	257
Table A.18	Plot arguments	257
Table A.19	Graphic parameters	258
Table A.20	Hypothesis tests and confidence intervals	258
Table A.21	Model fitting	258
Table A.22	<code>sixSigma</code> package data sets	259
Table A.23	<code>sixSigma</code> package functions	259
Table A.24	Cited packages	260

Acronyms

6 Ms	Six “ems”: Machines, Methods, Materials, Measurements, Mother Nature (environment), Manpower (people).
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
BB	Black Belt (role)
cdf	cumulative distribution function
CI	Confidence Interval
CL	Center Line
COQ	Cost of Quality
Cp	Process Capability Index
Cpk	Adjusted Process Capability Index
CRAN	Comprehensive R Archive Network
CTQ	Critical to Quality
DMAIC	Define, Measure, Analyze, Improve, Control
DFSS	Design for Six Sigma
DMADOV	Define, Measure, Analyze, Design, Optimize, Verify
DMADV	Define, Measure, Analyze, Design, Verify
DMEDI	Define, Measure, Explore, Develop, Implement
DoE	Design of Experiments
DPMO	Defects Per Million Opportunities
DPPM	Defective Parts Per Million
DPO	Defects Per Opportunity
DPU	Defects per Unit
EWMA	Exponentially Weighted Moving Average
FMECA	Failure Mode, Effect and Criticality Analysis
FTY	First Time Yield
GAM	Generalized Additive Model
GB	Green Belt (role)
GLM	Generalized Linear Model
GNU	GNU’s Not Unix! (a form of licencing software)
GUI	Graphical User Interface

HTML	HyperText Markup Language
IDOV	Identify, Design, Optimize, Validate
IQR	Interquartile Range
ISO	International Organization for Standardization
IT	Information Technologies
LCL	Lower Control Limit
LSL	Lower Specification Limits
LT	Long Term
MANOVA	Multivariate Analysis of Variance
MBB	Master Black Belt (role)
MDI	Multiple Document Interface
MSA	Measurement System Analysis
MSD	Mean Squared Deviation
NaN	Not a Number
NA	Not Available
NN	Neural Network
OC	Operating Characteristic
ODBC	Open DataBase Connection
PLS	Partial Least Squares
Q-Q	Quantile-Quantile
QC	Quality Control
R&R	Repeatability & Reproducibility
RNG	Random Number Generation
RPN	Risk Priority Number
RFT	Right First Time
RTY	Rolled Throughput Yield
RUMBA	Reasonable, Understandable, Measurable, Believable, Achievable
SD	Standard Deviation
SDI	Simple Document Interface
SDMAIC	Select, Define, Measure, Analyze, Improve, Control
SIPOC	Suppliers, Inputs, Process, Outputs, Customers
SMART	Specific, Measurable, Achievable, Relevant, Timed (SMART objectives)
SPC	Statistical Process Control
ST	Short Term
SVM	Support Vector Machine
UCL	Upper Control Limit
USL	Upper Specification Limit
VOC	Voice of the Customer
VOP	Voice of the Process
VSM	Value Stream Mapping
Z	Sigma score