Statistics for Industry, Technology, and Engineering

Series Editor

David Steinberg, Tel Aviv University, Tel Aviv, Israel

Editorial Board Members

V. Roshan Joseph, Georgia Institute of Technology, Atlanta, GA, USA

Ron S. Kenett, KPA Ltd. Raanana and Samuel Neaman Institute, Technion, Haifa, Israel

Christine Anderson-Cook, Los Alamos National Laboratory, Los Alamos, USA

Bradley Jones, SAS Institute, JMP Division, Cary, USA

Fugee Tsung, Hong Kong University of Science and Technology, Hong Kong, Hong Kong

The *Statistics for Industry, Technology, and Engineering* series will present up-to-date statistical ideas and methods that are relevant to researchers and accessible to an interdisciplinary audience: carefully organized authoritative presentations, numerous illustrative examples based on current practice, reliable methods, realistic data sets, and discussions of select new emerging methods and their application potential. Publications will appeal to a broad interdisciplinary readership including both researchers and practitioners in applied statistics, data science, industrial statistics, engineering statistics, quality control, manufacturing, applied reliability, and general quality improvement methods.

Principal Topic Areas:
* Quality Monitoring * Engineering Statistics * Data Analytics * Data Science * Time Series with Applications * Systems Analytics and Control * Stochastics and Simulation * Reliability * Risk Analysis * Uncertainty Quantification * Decision Theory * Survival Analysis * Prediction and Tolerance Analysis * Multivariate Statistical Methods * Nondestructive Testing * Accelerated Testing * Signal Processing * Experimental Design * Software Reliability * Neural Networks *

The series will include professional expository monographs, advanced textbooks, handbooks, general references, thematic compilations of applications/case studies, and carefully edited survey books.

Ron S. Kenett • Shelemyahu Zacks • Peter Gedeck

Modern Statistics

A Computer-Based Approach with Python

Birkhäuser

Ron S. Kenett
KPA Ltd. Raanana and Samuel Neaman Institute, Technion
Haifa, Israel

Shelemyahu Zacks
Mathematical Sciences
Binghamton University
Mc Lean, VA, USA

Peter Gedeck
Data Science
University of Virginia
Falls Church, VA, USA

ISSN 2662-5555 ISSN 2662-5563 (electronic)
Statistics for Industry, Technology, and Engineering
ISBN 978-3-031-07565-0 ISBN 978-3-031-07566-7 (eBook)
https://doi.org/10.1007/978-3-031-07566-7

Mathematics Subject Classification: 62E15, 62G30, 62M10, 62P30, 62P10, 97K40, 97K70, 97K80

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
This work is subject to copyright. All rights are solely and exclusively licensed 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.
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.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This book is published under the imprint Birkhäuser, www.birkhauser-science.com by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To my wife Sima, our children and their children: Yonatan, Alma, Tomer, Yadin, Aviv, Gili, Matan, Eden, and Ethan. RSK

To my wife Hanna, our sons Yuval and David, and their families with love. SZ

To Janet with love. PG

Preface

Statistics has developed by combining the needs of science, business, industry, and government. More recent development is connected with methods for generating insights from data, using statistical theory and delivery platforms. This integration is at the core of applied statistics and most of theoretical statistics.

Before the beginning of the twentieth century, statistics meant observed data and descriptive summary figures, such as means, variances, indices, etc., computed from data. With the introduction of the χ^2-test for goodness of fit by Karl Pearson (1900) and the t-test by Gosset (Student, 1908) for drawing inference on the mean of a normal population, statistics became a methodology of analyzing sample data to determine the validity of hypotheses about the source of the data (the population). Fisher (1922) laid the foundations for statistics as a discipline. He considered the object of statistical methods to be reducing data into the essential statistics, and he identified three problems that arise in doing so:

1. Specification-choosing the right mathematical model for a population
2. Estimation-methods to calculate, from a sample, estimates of the parameters of the hypothetical population
3. Distribution-properties of statistics derived from samples

Forty years later, Tukey (1962) envisioned a data-centric development of statistics, sketching the pathway to data science. Forty years after that, we entered the age of big data, data science, artificial intelligence, and machine learning. These new developments are built on the methods, applications, and experience of statisticians around the world.

The first two authors started collaborating on a book in the early 1990s. In 1998, we published with Duxbury Wadsworth *Modern Industrial Statistics: Design and Control of Quality and Reliability*. The book appeared in a Spanish edition (Estadística Industrial Moderna: Diseño y Control de Calidad y la Confiabilidad, Thomson International, 2000). An abbreviated edition was published as *Modern Statistics: A Computer based Approach* (Thomson Learning, 2001); this was followed by a Chinese edition (China Statistics Press, 2003) and a softcover edition, (Brooks/Cole, 2004). The book used QuickBasic, S-Plus, and MINITAB. In 2014

we published, with Wiley, an extended second edition titled *Modern Industrial Statistics: With Applications in R, MINITAB and JMP*. That book was translated into Vietnamese by the Vietnam Institute for Advanced Studies in Mathematics (VIASM, 2016). A third, expanded edition, was published by Wiley in 2021.

This book is about modern statistics with Python. It reflects many years of experience of the authors in doing research, teaching and applying statistics in science, healthcare, business, defense, and industry domains. The book invokes over 40 case studies and provides comprehensive Python applications. In 2019, there were 8.2 million developers in the world who code using Python which is considered the fastest-growing programming language. A special Python package, `mistat`, is available for download https://gedeck.github.io/mistat-code-solutions/ModernStatistics/. Everything in the book can be reproduced with `mistat`. We therefore provide, in this book, an integration of needs, methods, and delivery platform for a large audience and a wide range of applications.

Modern Statistics: A Computer-Based Approach with Python is a companion text to another book published by Springer titled: *Industrial Statistics: A Computer Based Approach with Python*. Both books include mutual cross references, but both books are stand-alone publications. This book can be used as textbook in a one semester or two semester course on modern statistics. The technical level of the presentation in both books can serve both undergraduate and graduate students. The example and case studies provide access to hands on teaching and learning. Every chapter includes exercises, data sets, and Python applications. These can be used in regular classroom setups, flipped classroom setups, and online or hybrid education programs. The companion text is focused on industrial statistics with special chapters on advanced process monitoring methods, cybermanufacturing, computer experiments, and Bayesian reliability. *Modern Statistics* is a foundational text and can be combined with any program requiring data analysis in its curriculum. This, for example, can be courses in data science, industrial statistics, physics, biology, chemistry, economics, psychology, social sciences, or any engineering discipline.

Modern Statistics: A Computer-Based Approach with Python includes eight chapters. Chapter 1 is on analyzing variability with descriptive statistics. Chapter 2 is on probability models and distribution functions. Chapter 3 introduces statistical inference and bootstrapping. Chapter 4 is on variability in several dimensions and regression models. Chapter 5 covers sampling for estimation of finite population quantities, a common situation when one wants to infer on a population from a sample. Chapter 6 is dedicated to time series analysis and prediction. Chapters 7 and 8 are about modern data analytic methods.

Industrial Statistics: A Computer-Based Approach with Python contains 11 chapters: Chapter 1— Introduction to Industrial Statistics, Chapter 2—Basic Tools and Principles of Process Control, Chapter 3—Advanced Methods of Statistical Process Control, Chapter 4—Multivariate Statistical Process Control, Chapter 5—Classical Design and Analysis of Experiments, Chapter 6—Quality by Design, Chapter 7—Computer Experiments, Chapter 8—Cybermanufacturing and Digital Twins, Chapter 9—Reliability Analysis, Chapter 10—Bayesian Reliability Estima-

tion and Prediction, and Chapter 11—Sampling Plans for Batch and Sequential Inspection. This second book is focused on industrial statistics with applications to monitoring, diagnostics, prognostic, and prescriptive analytics. It can be used as a stand-alone book, or in conjunction with *Modern Statistics*. Both books include solution manuals to exercises listed at the end of each chapter. This was designed to support self-learning as well as instructor led courses.

We made every possible effort to ensure the calculations are correct and the text is clear. However, should errors have skipped to the printed version, we would appreciate feedback from readers noticing these. In general, any feedback will be much appreciated.

Finally, we would like to thank the team at Springer Birkhäuser, including Dana Knowles and Christopher Tominich. They made everything in the publication process look easy.

Ra'anana, Israel — Ron S. Kenett
McLean, VA, USA — Shelemyahu Zacks
Falls Church, VA, USA — Peter Gedeck
April 2022

Contents

1 Analyzing Variability: Descriptive Statistics 1
1.1 Random Phenomena and the Structure of Observations 1
1.2 Accuracy and Precision of Measurements 6
1.3 The Population and the Sample 8
1.4 Descriptive Analysis of Sample Values 9
1.4.1 Frequency Distributions of Discrete Random Variables ... 9
1.4.2 Frequency Distributions of Continuous Random Variables 14
1.4.3 Statistics of the Ordered Sample 17
1.4.4 Statistics of Location and Dispersion 19
1.5 Prediction Intervals 23
1.6 Additional Techniques of Exploratory Data Analysis 25
1.6.1 Density Plots 25
1.6.2 Box and Whiskers Plots 27
1.6.3 Quantile Plots 29
1.6.4 Stem-and-Leaf Diagrams 30
1.6.5 Robust Statistics for Location and Dispersion 31
1.7 Chapter Highlights 34
1.8 Exercises 34

2 Probability Models and Distribution Functions 39
2.1 Basic Probability 39
2.1.1 Events and Sample Spaces: Formal Presentation of Random Measurements 39
2.1.2 Basic Rules of Operations with Events: Unions and Intersections 41
2.1.3 Probabilities of Events 44
2.1.4 Probability Functions for Random Sampling 46
2.1.5 Conditional Probabilities and Independence of Events 49
2.1.6 Bayes' Theorem and Its Application 51
2.2 Random Variables and Their Distributions 54

2.2.1 Discrete and Continuous Distributions 55
2.2.1.1 Discrete Random Variables 55
2.2.1.2 Continuous Random Variables 56
2.2.2 Expected Values and Moments of Distributions 59
2.2.3 The Standard Deviation, Quantiles, Measures of Skewness, and Kurtosis 62
2.2.4 Moment Generating Functions 65
2.3 Families of Discrete Distribution 66
2.3.1 The Binomial Distribution 66
2.3.2 The Hypergeometric Distribution 69
2.3.3 The Poisson Distribution 72
2.3.4 The Geometric and Negative Binomial Distributions 74
2.4 Continuous Distributions 78
2.4.1 The Uniform Distribution on the Interval (a, b), $a < b$ 78
2.4.2 The Normal and Log-Normal Distributions 79
2.4.2.1 The Normal Distribution 79
2.4.2.2 The Log-Normal Distribution 84
2.4.3 The Exponential Distribution 85
2.4.4 The Gamma and Weibull Distributions 88
2.4.5 The Beta Distributions 92
2.5 Joint, Marginal, and Conditional Distributions 93
2.5.1 Joint and Marginal Distributions 93
2.5.2 Covariance and Correlation 96
2.5.3 Conditional Distributions 99
2.6 Some Multivariate Distributions 102
2.6.1 The Multinomial Distribution 102
2.6.2 The Multi-Hypergeometric Distribution 104
2.6.3 The Bivariate Normal Distribution 105
2.7 Distribution of Order Statistics 108
2.8 Linear Combinations of Random Variables 111
2.9 Large Sample Approximations 117
2.9.1 The Law of Large Numbers 117
2.9.2 The Central Limit Theorem 117
2.9.3 Some Normal Approximations 119
2.10 Additional Distributions of Statistics of Normal Samples 120
2.10.1 Distribution of the Sample Variance 121
2.10.2 The "Student" t-Statistic 122
2.10.3 Distribution of the Variance Ratio 123
2.11 Chapter Highlights 125
2.12 Exercises 126

3 Statistical Inference and Bootstrapping 139
3.1 Sampling Characteristics of Estimators 139
3.2 Some Methods of Point Estimation 141

3.2.1 Moment Equation Estimators 142
3.2.2 The Method of Least Squares 144
3.2.3 Maximum Likelihood Estimators 146
3.3 Comparison of Sample Estimates 149
3.3.1 Basic Concepts 149
3.3.2 Some Common One-Sample Tests of Hypotheses 152
3.3.2.1 The Z-Test: Testing the Mean of a Normal Distribution, σ^2 Known 152
3.3.2.2 The t-Test: Testing the Mean of a Normal Distribution, σ^2 Unknown 155
3.3.2.3 The Chi-Squared Test: Testing the Variance of a Normal Distribution 156
3.3.2.4 Testing Hypotheses About the Success Probability, p, in Binomial Trials 158
3.4 Confidence Intervals 160
3.4.1 Confidence Intervals for μ; σ Known 161
3.4.2 Confidence Intervals for μ; σ Unknown 162
3.4.3 Confidence Intervals for σ^2 162
3.4.4 Confidence Intervals for p 163
3.5 Tolerance Intervals 166
3.5.1 Tolerance Intervals for the Normal Distributions 166
3.6 Testing for Normality with Probability Plots 169
3.7 Tests of Goodness of Fit 173
3.7.1 The Chi-Square Test (Large Samples) 173
3.7.2 The Kolmogorov-Smirnov Test 175
3.8 Bayesian Decision Procedures 176
3.8.1 Prior and Posterior Distributions 177
3.8.2 Bayesian Testing and Estimation 181
3.8.2.1 Bayesian Testing 181
3.8.2.2 Bayesian Estimation 184
3.8.3 Credibility Intervals for Real Parameters 185
3.9 Random Sampling from Reference Distributions 186
3.10 Bootstrap Sampling 189
3.10.1 The Bootstrap Method 189
3.10.2 Examining the Bootstrap Method 190
3.10.3 Harnessing the Bootstrap Method 192
3.11 Bootstrap Testing of Hypotheses 192
3.11.1 Bootstrap Testing and Confidence Intervals for the Mean 192
3.11.2 Studentized Test for the Mean 193
3.11.3 Studentized Test for the Difference of Two Means 195
3.11.4 Bootstrap Tests and Confidence Intervals for the Variance 197
3.11.5 Comparing Statistics of Several Samples 199
3.11.5.1 Comparing Variances of Several Samples 200

3.11.5.2 Comparing Several Means: The One-Way Analysis of Variance 201
3.12 Bootstrap Tolerance Intervals 204
3.12.1 Bootstrap Tolerance Intervals for Bernoulli Samples 204
3.12.2 Tolerance Interval for Continuous Variables 205
3.12.3 Distribution-Free Tolerance Intervals 206
3.13 Non-Parametric Tests 208
3.13.1 The Sign Test 208
3.13.2 The Randomization Test 210
3.13.3 The Wilcoxon Signed-Rank Test 211
3.14 Chapter Highlights 214
3.15 Exercises 215

4 Variability in Several Dimensions and Regression Models 225
4.1 Graphical Display and Analysis 226
4.1.1 Scatterplots 226
4.1.2 Multiple Boxplots 229
4.2 Frequency Distributions in Several Dimensions 230
4.2.1 Bivariate Joint Frequency Distributions 231
4.2.2 Conditional Distributions 234
4.3 Correlation and Regression Analysis 235
4.3.1 Covariances and Correlations 236
4.3.2 Fitting Simple Regression Lines to Data 237
4.3.2.1 The Least Squares Method 239
4.3.2.2 Regression and Prediction Intervals 243
4.4 Multiple Regression 245
4.4.1 Regression on Two Variables 246
4.4.2 Partial Regression and Correlation 251
4.4.3 Multiple Linear Regression 254
4.4.4 Partial-F Tests and the Sequential SS 260
4.4.5 Model Construction: Step-Wise Regression 263
4.4.6 Regression Diagnostics 265
4.5 Quantal Response Analysis: Logistic Regression 268
4.6 The Analysis of Variance: The Comparison of Means 271
4.6.1 The Statistical Model 271
4.6.2 The One-Way Analysis of Variance (ANOVA) 271
4.7 Simultaneous Confidence Intervals: Multiple Comparisons 275
4.8 Contingency Tables 279
4.8.1 The Structure of Contingency Tables 279
4.8.2 Indices of association for contingency tables 282
4.8.2.1 Two Interval-Scaled Variables 282
4.8.2.2 Indices of Association for Categorical Variables 284
4.9 Categorical Data Analysis 288
4.9.1 Comparison of Binomial Experiments 288

4.10 Chapter Highlights 290
4.11 Exercises 291

5 Sampling for Estimation of Finite Population Quantities 299
5.1 Sampling and the Estimation Problem 299
5.1.1 Basic Definitions 299
5.1.2 Drawing a Random Sample from a Finite Population 301
5.1.3 Sample Estimates of Population Quantities and Their Sampling Distribution 302
5.2 Estimation with Simple Random Samples 305
5.2.1 Properties of $\bar{X}_n$ and S_n^2 Under RSWR 306
5.2.2 Properties of $\bar{X}_n$ and S_n^2 Under RSWOR 310
5.3 Estimating the Mean with Stratified RSWOR 314
5.4 Proportional and Optimal Allocation 316
5.5 Prediction Models with Known Covariates 320
5.6 Chapter Highlights 324
5.7 Exercises 325

6 Time Series Analysis and Prediction 329
6.1 The Components of a Time Series 330
6.1.1 The Trend and Covariances 330
6.1.2 Analyzing Time Series with Python 331
6.2 Covariance Stationary Time Series 336
6.2.1 Moving Averages 337
6.2.2 Auto-Regressive Time Series 338
6.2.3 Auto-Regressive Moving Average Time Series 343
6.2.4 Integrated Auto-Regressive Moving Average Time Series 344
6.2.5 Applications with Python 345
6.3 Linear Predictors for Covariance Stationary Time Series 346
6.3.1 Optimal Linear Predictors 346
6.4 Predictors for Non-stationary Time Series 349
6.4.1 Quadratic LSE Predictors 349
6.4.2 Moving Average Smoothing Predictors 351
6.5 Dynamic Linear Models 352
6.5.1 Some Special Cases 353
6.5.1.1 The Normal Random Walk 353
6.5.1.2 Dynamic Linear Model With Linear Growth ... 354
6.5.1.3 Dynamic Linear Model for ARMA(p,q) 355
6.6 Chapter Highlights 358
6.7 Exercises 359

7 Modern Analytic Methods: Part I 361
7.1 Introduction to Computer Age Statistics 361
7.2 Data Preparation 362
7.3 The Information Quality Framework 363

7.4 Determining Model Performance 364
7.5 Decision Trees 368
7.6 Ensemble Models 376
7.7 Naïve Bayes Classifier 378
7.8 Neural Networks 381
7.9 Clustering Methods 386
7.9.1 Hierarchical Clustering 386
7.9.2 K-Means Clustering 389
7.9.3 Cluster Number Selection 390
7.10 Chapter Highlights 392
7.11 Exercises 392

8 Modern Analytic Methods: Part II 395
8.1 Functional Data Analysis 395
8.2 Text Analytics 401
8.3 Bayesian Networks 405
8.4 Causality Models 411
8.5 Chapter Highlights 416
8.6 Exercises 417

A Introduction to Python 421
A.1 List, Set, and Dictionary Comprehensions 421
A.2 Pandas Data Frames 422
A.3 Data Visualization Using `Pandas` and `Matplotlib` 423

B List of Python Packages 425

C Code Repository and Solution Manual 427

Bibliography 429

Index 433

Industrial Statistics: A Computer-Based Approach with Python (Companion volume)

1 The Role of Statistical Methods in Modern Industry
1.1 The Evolution of Industry
1.2 The Evolution of Quality
1.3 Industry 4.0 Characteristics
1.4 The Digital Twin
1.5 Chapter Highlights
1.6 Exercises

2 Basic Tools and Principles of Process Control
2.1 Basic Concepts of Statistical Process Control
2.2 Driving a Process with Control Charts
2.3 Setting Up a Control Chart: Process Capability Studies
2.4 Process Capability Indices
2.5 Seven Tools for Process Control and Process Improvement
2.6 Statistical Analysis of Pareto Charts
2.7 The Shewhart Control Charts
2.8 Process analysis with data segments
2.9 Chapter Highlights
2.10 Exercises

3 Advanced Methods of Statistical Process Control
3.1 Tests of Randomness
3.2 Modified Shewhart Control Charts for $\bar{X}$
3.3 The Size and Frequency of Sampling for Shewhart Control Charts
3.4 Cumulative Sum Control Charts
3.5 Bayesian Detection
3.6 Process Tracking
3.7 Automatic Process Control
3.8 Chapter Highlights
3.9 Exercises

4 Multivariate Statistical Process Control
4.1 Introduction
4.2 A Review Multivariate Data Analysis
4.3 Multivariate Process Capability Indices
4.4 Advanced Applications of Multivariate Control Charts
4.5 Multivariate Tolerance Specifications
4.6 Tracking structural changes
4.7 Chapter Highlights
4.8 Exercises

5 Classical Design and Analysis of Experiments
5.1 Basic Steps and Guiding Principles
5.2 Blocking and Randomization
5.3 Additive and Non-Additive Linear Models
5.4 The Analysis of Randomized Complete Block Designs
5.5 Balanced Incomplete Block Designs
5.6 Latin Square Design
5.7 Full Factorial Experiments
5.8 Blocking and Fractional Replications of 2^m Factorial Designs
5.9 Exploration of Response Surfaces
5.10 Evaluating Designed Experiments
5.11 Chapter Highlights
5.12 Exercises

6 Quality by Design
6.1 Off-Line Quality Control, Parameter Design and The Taguchi Method
6.2 The Effects of Non-Linearity
6.3 Taguchi's Designs
6.4 Quality by Design in the Pharmaceutical Industry
6.5 Tolerance Designs
6.6 Case Studies
6.7 Chapter Highlights
6.8 Exercises

7 Computer Experiments
7.1 Introduction to Computer Experiments
7.2 Designing Computer Experiments
7.3 Analyzing Computer Experiments
7.4 Stochastic Emulators
7.5 Integrating Physical and Computer Experiments
7.6 Simulation of Random Variables
7.7 Chapter Highlights
7.8 Exercises

8 Cybermanufacturing and Digital Twins
8.1 Introduction to Cybermanufacturing
8.2 Cybermanufacturing Analytics
8.3 Information Quality in Cybermanufacturing
8.4 Modeling in Cybermanufacturing
8.5 Computational pipelines
8.6 Digital Twins
8.7 Chapter Highlights
8.8 Exercises

9 Reliability Analysis
9.1 Basic Notions
9.2 System Reliability
9.3 Availability of Repairable Systems
9.4 Types of Observations on TTF
9.5 Graphical Analysis of Life Data
9.6 Non-Parametric Estimation of Reliability
9.7 Estimation of Life Characteristics
9.8 Reliability Demonstration
9.9 Accelerated Life Testing
9.10 Burn-In Procedures
9.11 Chapter Highlights
9.12 Exercises

10 Bayesian Reliability Estimation and Prediction
10.1 Prior and Posterior Distributions
10.2 Loss Functions and Bayes Estimators
10.3 Bayesian Credibility and Prediction Intervals
10.4 Credibility Intervals for the Asymptotic Availability of Repairable Systems: The Exponential Case
10.5 Empirical Bayes Method
10.6 Chapter Highlights
10.7 Exercises

11 Sampling Plans for Batch and Sequential Inspection
11.1 General Discussion
11.2 Single-Stage Sampling Plans for Attributes
11.3 Approximate Determination of the Sampling Plan
11.4 Double-Sampling Plans for Attributes
11.5 Sequential Sampling and A/B testing
11.6 Acceptance Sampling Plans for Variables
11.7 Rectifying Inspection of Lots
11.8 National and International Standards
11.9 Skip-Lot Sampling Plans for Attributes
11.10 The Deming Inspection Criterion
11.11 Published Tables for Acceptance Sampling

11.12 Sequential Reliability Testing
11.13 Chapter Highlights
11.14 Exercises
References

A List of Python packages

List of Abbreviations

AIC	Akaike information criteria
ANOVA	Analysis of variance
ANSI	American National Standard Institute
AOQ	Average outgoing quality
AOQL	Average outgoing quality limit
AQL	Acceptable quality level
ARIMA	Autoregressive integrated moving average
ARL	Average run length
ASN	Average sample number
ASQ	American Society for Quality
ATE	Average treatment effect
ATI	Average total inspection
BECM	Bayes estimation of the current mean
BI	Business intelligence
BIBD	Balanced incomplete block design
BIC	Bayesian information criteria
BN	Bayesian network
BP	Bootstrap population
c.d.f.	Cumulative distribution function
CAD	Computer-aided design
CADD	Computer-aided drawing and drafting
CAM	Computer-aided manufacturing
CART	Classification and regression trees
CBD	Complete block design
CED	Conditional expected delay
cGMP	Current good manufacturing practices
CHAID	Chi-square automatic interaction detector
CIM	Computer integrated manufacturing
CLT	Central limit theorem
CMM	Coordinate measurement machines
CMMI	Capability maturity model integrated

CNC	Computerized numerically controlled
CPA	Circuit pack assemblies
CQA	Critical quality attribute
CUSUM	Cumulative sum
DACE	Design and analysis of computer experiments
DAG	Directed acyclic graph
DFIT	Difference in fits distance
DLM	Dynamic linear model
DoE	Design of experiments
DTM	Document term matrix
EBD	Empirical bootstrap distribution
ETL	Extract-transform-load
EWMA	Exponentially weighted moving average
FDA	Food and Drug Administration
FDA	Functional data analysis
FPCA	Functional principal component analysis
FPM	Failures per million
GFS	Google file system
GRR	Gage repeatability and reproducibility
HPD	Highest posterior density
HPLC	High-performance liquid chromatography
i.i.d.	Independent and identically distributed
IDF	Inverse document frequency
InfoQ	Information quality
IPO	Initial public offering
IPS	Inline process control
IQR	Inter quartile range
ISC	Short circuit current of solar cells (in Ampere)
KS	Kolmogorov-Smirnov test
LCL	Lower control limit
LLN	Law of large numbers
LQL	Limiting quality level
LSA	Latent semantic analysis
LSL	Lower specification limit
LTPD	Lot tolerance percent defective
LWL	Lower warning limit
MAE	Mean absolute error
m.g.f.	Moment generating function
MLE	Maximum likelihood estimator
MSD	Mean squared deviation
MSE	Mean squared error
MTBF	Mean time between failures
MTTF	Mean time to failure
NID	Normal independently distributed
OAB	One-armed bandit

OC	Operating characteristic
p.d.f.	Probability density function
PCA	Principal component analysis
PERT	Project evaluation and review technique
PFA	Probability of false alarm
PL	Product limit estimator
PPM	Defects in parts per million
PSE	Practical statistical efficiency
QbD	Quality by design
QMP	Quality measurement plan
QQ-Plot	Quantile vs. quantile plot
RCBD	Randomized complete block design
Regex	Regularized expression
RMSE	Root mean squared error
RSWOR	Random sample without replacement
RSWR	Random sample with replacement
SE	Standard error
SL	Skip lot
SLOC	Source lines of code
SLSP	Skip lot sampling plans
SPC	Statistical process control
SPRT	Sequential probability ratio test
SR	Shiryaev Roberts
SSE	Sum of squares of errors
SSR	Sum of squares around the regression model
SST	Total sum of squares
STD	Standard deviation
SVD	Singular value decomposition
TAB	Two-armed bandit
TF	Term frequency
TTC	Time till censoring
TTF	Time till failure
TTR	Time till repair
TTT	Total time on test
UCL	Upper control limit
USL	Upper specification limit
UWL	Upper warning limit
WSP	Wave soldering process