Statistics for Industry, Technology, and Engineering

Ron S. Kenett
Shelemyahu Zacks
Peter Gedeck

Modern Statistics

A Computer-Based Approach with Python

MOREMEDIA

Birkhäuser

Statistics for Industry, Technology, and Engineering

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

This work contains media enhancements, which are displayed with a "play" icon. Material in the print book can be viewed on a mobile device by downloading the Springer Nature "More Media" app available in the major app stores. The media enhancements in the online version of the work can be accessed directly by authorized users.

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

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