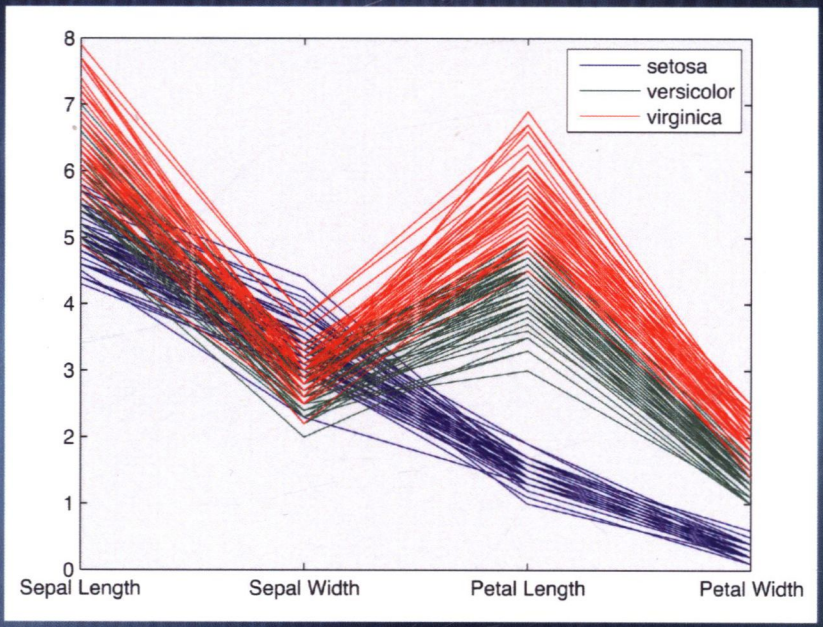
Computer Science and Data Analysis Series

STATISTICS IN MATLAB®

A PRIMER



WENDY L. MARTINEZ
MOONJUNG CHO

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Statistics

Computer Science and Data Analysis Series

Fulfilling the need for a practical user's guide, **Statistics in MATLAB: A Primer** provides an accessible introduction to the latest version of MATLAB® and its extensive functionality for statistics. Assuming a basic knowledge of statistics and probability as well as a fundamental understanding of linear algebra concepts, this book:

- Covers capabilities in the main MATLAB package, the Statistics Toolbox, and the student version of MATLAB
- Presents examples of how MATLAB can be used to analyze data
- Offers access to a companion website with data sets and additional examples
- Contains figures and visual aids to assist in application of the software
- Explains how to determine what method should be used for analysis

Statistics in MATLAB: A Primer is an ideal reference for undergraduate and graduate students in engineering, mathematics, statistics, economics, biostatistics, and computer science. It is also appropriate for a diverse professional market, making it a valuable addition to the libraries of researchers in statistics, computer science, data mining, machine learning, image analysis, signal processing, and engineering.

Authors: Wendy L. Martinez & MoonJung Cho

ISBN 978-1-138-46931-0

9 781138 469310

CRC Press
Taylor & Francis Group
an informa business
www.crcpress.com


CRC

Computer Science and Data Analysis Series

STATISTICS IN MATLAB®
A PRIMER

WENDY L. MARTINEZ
BUREAU OF LABOR STATISTICS
WASHINGTON, D.C., USA

MOONJUNG CHO
BUREAU OF LABOR STATISTICS
WASHINGTON, D.C., USA

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

First issued in hardback 2018

ISBN 13: 978-1-138-46931-0 (hbk)
ISBN 13: 978-1-4665-9656-6 (pbk)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

STATISTICS IN MATLAB®
A PRIMER

Chapman & Hall/CRC

Computer Science and Data Analysis Series

The interface between the computer and statistical sciences is increasing, as each discipline seeks to harness the power and resources of the other. This series aims to foster the integration between the computer sciences and statistical, numerical, and probabilistic methods by publishing a broad range of reference works, textbooks, and handbooks.

Published Titles

Semisupervised Learning for Computational Linguistics
Steven Abney

Visualization and Verbalization of Data
Jörg Blasius and Michael Greenacre

Design and Modeling for Computer Experiments
Kai-Tai Fang, Runze Li, and Agus Sudjianto

Microarray Image Analysis: An Algorithmic Approach
Karl Fraser, Zidong Wang, and Xiaohui Liu

R Programming for Bioinformatics
Robert Gentleman

Exploratory Multivariate Analysis by Example Using R
François Husson, Sébastien Lê, and Jérôme Pagès

Bayesian Artificial Intelligence, Second Edition
Kevin B. Korb and Ann E. Nicholson

Published Titles cont.

Computational Statistics Handbook with MATLAB®, Second Edition
Wendy L. Martinez and Angel R. Martinez

Exploratory Data Analysis with MATLAB®, Second Edition
Wendy L. Martinez, Angel R. Martinez, and Jeffrey L. Solka

Statistics in MATLAB®: A Primer
Wendy L. Martinez and MoonJung Cho

Clustering for Data Mining: A Data Recovery Approach, Second Edition
Boris Mirkin

Introduction to Machine Learning and Bioinformatics
Sushmita Mitra, Sujay Datta, Theodore Perkins, and George Michailidis

Introduction to Data Technologies
Paul Murrell

R Graphics
Paul Murrell

Correspondence Analysis and Data Coding with Java and R
Fionn Murtagh

Pattern Recognition Algorithms for Data Mining
Sankar K. Pal and Pabitra Mitra

Statistical Computing with R
Maria L. Rizzo

Statistical Learning and Data Science
Mireille Gettler Summa, Léon Bottou, Bernard Goldfarb, Fionn Murtagh, Catherine Pardoux, and Myriam Touati

Foundations of Statistical Algorithms: With References to R Packages
Claus Weihs, Olaf Mersmann, and Uwe Ligges

Wendy dedicates this book to her parents
who started her on this path:

Shirley and Glenn Cukr

MoonJung dedicates this book to her children:

Catherine and Ted

Preface

The functionality in MATLAB® for statistical data analysis has improved and expanded in the past several years, with major changes made in 2012 (version 8). Additionally, MATLAB is frequently used in academia to teach statistics, engineering, mathematics, and data analysis courses. Thus, we felt that a book that provides an overview or introduction to the extensive functionality available in MATLAB would be useful to a wide audience.

The main MATLAB software includes many basic functions for statistical visualization and data analysis. The MathWorks, Inc. Statistics Toolbox extends these basic capabilities by including additional specialized functions. The Statistics Toolbox can be purchased separately. MathWorks also has a Student Version of MATLAB that includes the Statistics Toolbox, as well as many other toolboxes.

One should have the Statistics Toolbox to get the most from this book. However, we include enough content to help those who do not have the extra MATLAB functionality. We have been careful to note where to find the functions—in the base MATLAB or the Statistics Toolbox. If the reader is ever confused about where the function comes from, type

```
which functionname
```

at the command line, and the location of the function file will be displayed.

For example, **normpdf** is a function in the Statistics Toolbox. We get the following, when we type **which normpdf**:

```
C:\MATLAB2013a\toolbox\stats\stats\normpdf.m
```

This is under the directory **~\toolbox\stats**, indicating the function is part of the Statistics Toolbox. A function that is in base MATLAB will have a directory **~\toolbox\matlab**.

It took over a year to write this book and several versions of MATLAB. We started with MATLAB R2013a and finished with MATLAB 2014a. All functions should work with R2013, with a few exceptions. These are found mostly in Chapter 8. We recommend that readers investigate the changes in MATLAB by looking at the release notes that are available in the MATLAB documentation or try to have the latest version installed.

There is a companion website, where the reader can find the data sets, M-files with the code from the book, and additional examples. This website is

```
www.pi-sigma.info
```

You can also download the files from the book website at CRC Press.

For the most part, we assume the reader has a basic knowledge of statistics and probability. The focus of this book is on how to use the statistics capabilities in MATLAB, not on the theory and use of statistics. However, we do include definitions and formulas in certain areas in order to aid the understanding of the MATLAB functions.

The reader should also know some basic concepts of linear algebra. This includes the definitions of vectors, matrices and operations such as adding vectors, multiplying matrices, taking transposes, etc.

Please note that this book is an *introduction*. It is not meant to cover all aspects of statistical analysis nor all of the functions available for statistics in MATLAB. Readers should always refer to the help files and other documentation provided with MATLAB to get the full story. We provide information on how to access the documentation in the first chapter.

We would like to acknowledge the invaluable help of the reviewers: Tom Lane, John Eltinge, Terrance Savitsky, Eungchun Cho, Ted Cho, and Angel Martinez. Their many helpful comments and suggestions resulted in a better book. Any shortcomings are the sole responsibility of the authors. We greatly appreciate the help and patience of those at CRC Press: David Grubbs, Jessica Vakili, Robin Starkes, and Kevin Craig. Finally, we are indebted to Naomi Fernandes, Paul Pilotte, and Tom Lane at The MathWorks, Inc. for their special assistance with MATLAB.

Disclaimers

1. Any MATLAB programs and data sets that are included with the book are provided in good faith. The authors, publishers, or distributors do not guarantee their accuracy and are not responsible for the consequences of their use.
2. MATLAB® and Simulink® are trademarks of the MathWorks, Inc. and are used with permission. The MathWorks does not warrant the accuracy of the text or the exercises in this book. This book's use or discussion of MATLAB® and Simulink® software or related products does not constitute endorsement or sponsorship by the MathWorks of a particular pedagogical approach or particular use of the MATLAB® and Simulink® software.
3. The views expressed in this book are those of the authors and do not necessarily represent the views of the United States Department of Labor or its components.

Wendy Martinez
MoonJung Cho

Table of Contents