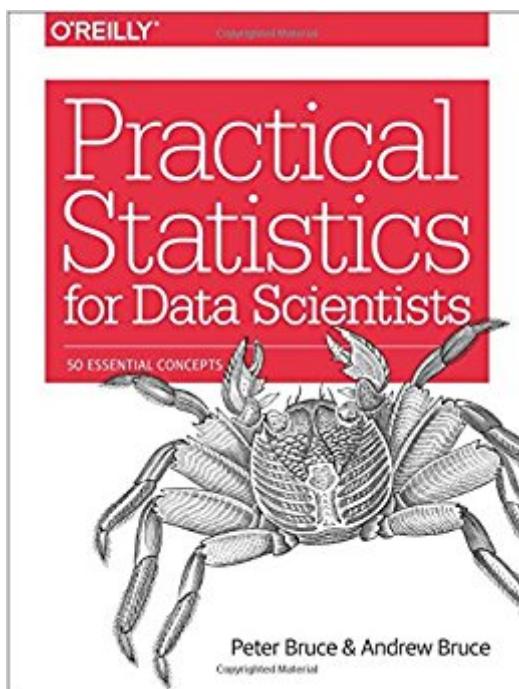The book was found

Practical Statistics For Data Scientists: 50 Essential Concepts

Synopsis

Statistical methods are a key part of of data science, yet very few data scientists have any formal statistics training. Courses and books on basic statistics rarely cover the topic from a data science perspective. This practical guide explains how to apply various statistical methods to data science, tells you how to avoid their misuse, and gives you advice on what's important and what's not.Many data science resources incorporate statistical methods but lack a deeper statistical perspective. If youÃ¢â¬â„¢re familiar with the R programming language, and have some exposure to statistics, this quick reference bridges the gap in an accessible, readable format.With this book, youÃ¢â¬â„¢ll learn:Why exploratory data analysis is a key preliminary step in data scienceHow random sampling can reduce bias and yield a higher quality dataset, even with big dataHow the principles of experimental design yield definitive answers to questionsHow to use regression to estimate outcomes and detect anomaliesKey classification techniques for predicting which categories a record belongs toStatistical machine learning methods that Ã¢â¬Å"learnÃ¢â¬Å• from dataUnsupervised learning methods for extracting meaning from unlabeled data

Book Information

Paperback: 318 pages

Publisher: O'Reilly Media; 1 edition (May 28, 2017)

Language: English

ISBN-10: 1491952962

ISBN-13: 978-1491952962

Product Dimensions: 6.9 x 0.6 x 9.1 inches

Shipping Weight: 1.2 pounds (View shipping rates and policies)

Average Customer Review: 3.8 out of 5 stars 14 customer reviews

Best Sellers Rank: #20,126 in Books (See Top 100 in Books) #2 inÃ Â Books > Computers & Technology > Databases & Big Data > Data Warehousing #17 inÃ Â Books > Computers & Technology > Software > Mathematical & Statistical #23 inÃ Â Books > Computers & Technology > Databases & Big Data > Data Processing

Customer Reviews

View larger About this Book Data science is a fusion of multiple disciplines, including statistics, computer science, information technology and domain specific fields. As a result, a several different terms could be used to reference a given concept. Key terms and their synonyms will be highlighted throughout the book in a sidebar within the text. This book is aimed at the data scientist with some

familiarity with the R programming language, and with some prior (perhaps spotty or ephemeral) exposure to statistics. Both of us came to the world of data science from the world of statistics, and have some appreciation of the contribution that statistics can make to the art of data science. At the same time, we are well aware of the limitations of traditional statistics instruction: statistics as a disciple is a century and a half old, and most statistics textbooks and courses are laden with the momentum and inertia worthy of an ocean liner.   Two goals underlie this book: To lay out, in digestible, navigable and easily referenced form, key concepts from statistics that are relevant to data science.   To explain which concepts are important and useful from a data science perspective, which are less so, and why.

50 Essential Concepts

Peter Bruce founded and grew the Institute for Statistics Education at Statistics.com, which now offers about 100 courses in statistics, roughly a third of which are aimed at the data scientist. In recruiting top authors as instructors and forging a marketing strategy to reach professional data scientists, Peter has developed both a broad view of the target market, and his own expertise to reach it.Andrew Bruce has over 30 years of experience in statistics and data science in academia, government and business. He has a Ph.D. in statistics from the University of Washington and published numerous papers in refereed journals. He has developed statistical-based solutions to a wide range of problems faced by a variety of industries, from established financial firms to internet startups, and offers a deep understanding the practice of data science.

This book is well written and packs a substantial amount of information into a small number of pages. It is best used to get a survey and overview of many of the facets of the domain of data science. This book will not teach you anything in enough depth to actually execute it well ÃƒÂ¢Ã â ¬Ã â • it will teach you just enough to be dangerous and not realize when you've gone off the rails. I recommend it for managers who may never go into technical depth, for people considering whether or not they are interested in data science, or as a preview book to create a framework from which to hang more detailed understanding. Although this is an introductory book, it assumes you can already program in R. If you can't, either accept that you won't be able to follow the specifics of the examples, or read The Art of R Programming and/or R for Data Science.I dislike that the authors make a number of categorical statements of the form "Data Scientists do this" or "Data Scientists don't need that". I disagree with many of these assertions and I think they have

taken a definition of "data science" which is narrower than the prevailing consensus in the industry.This book has some errors (see, for example, the confusion matrix on page 196) but overall the accuracy is above average relative to recent norms.As other reviewers have noted, the author's github repository for the book is currently empty. If that's important to you, check it under "andrewgbruce" on github and make sure it's been updated before you buy the book.

There's always that one person who is unsatisfied, but it sure as hell isn't me, because I knew what this book was going to be like the moment I saw how many pages it was going to have & how the early release version looked. I still preordered a hard copy (for sharing) & a digital copy (for carrying), because I knew this was kind of what type of book I was looking for & then some.The concepts are not astronomically explained, but with just enough depth that I can also individually explain to people what they are. What really stands out for me so far is after each or so concept, there is a section labeled as further reading (well, in the digital copy) that is usually at the end of the book altogether & I found myself realizing I have a lot of those books so the authors really know where to look & guide those who wanted more depth.Yeah yeah yeah, the codes are missing (as of mid-June 2017) but if you really understood / know which packages to use, you wouldn't need the code. The first half of the book are two three liners of code concepts anyways; it's the explanations that matter the most. The second half of the book is the good part, which separates a white hat statistician from a grey hat data scientist, which is exactly what I wanted in a

Information seems plainly written and relevant. No link to datasets makes the "practical" code portion of the book unusable. Will happily update my review when the datasets are released.EDIT:Ok the datasets are up. There is a short R script to run to download the data, it will require some small modifications to get it working correctly.You need to create a folder named "data".and I changed the second line in the script from:PSDS_PATH

Excellent introductory text for a comprehensive overview of statistics! The github repository augments the content very well and provides added value for the statistical topics covered in the book. Both of the Bruce brothers are statistical gurus and this fact is evident in the writing, which is both informative and witty. Peter is the president of Statistics.com and is well-versed in providing statistical instruction to students of all ages and levels. He is also a proponent of resampling and one of the developers of the excellent Resampling Stats software package for Excel.It is true that the textbook does not provide in-depth coverage for all topics, but I don't think that was the intent of

the authors. However, the text DOES provide an excellent introduction to topics relevant to students and data scientists. After reading the text and working through the examples, you will be equipped to further your knowledge in whichever topic you require for you data analysis task.Highly recommended!

Awaiting the GitHub link associated with the book to be updated. Currently the link returns empty page. Thanks !

Practical Statistics for Data Scientists: 50 Essential Concepts Data Analytics: What Every Business Must Know About Big Data And Data Science (Data Analytics for Business, Predictive Analysis, Big Data Book 1) Data Analytics: Applicable Data Analysis to Advance Any Business Using the Power of Data Driven Analytics (Big Data Analytics, Data Science, Business Intelligence Book 6) Big Data For Business: Your Comprehensive Guide to Understand Data Science, Data Analytics and Data Mining to Boost More Growth and Improve Business - Data Analytics Book, Series 2 Statistics for People Who (Think They) Hate Statistics (Salkind, Statistics for People Who(Think They Hate Statistics(Without CD)) Analytics: Data Science, Data Analysis and Predictive Analytics for Business (Algorithms, Business Intelligence, Statistical Analysis, Decision Analysis, Business Analytics, Data Mining, Big Data) Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data (Princeton Series in Modern Observational Astronomy) Statistics and Data Analysis for Financial Engineering: with R examples (Springer Texts in Statistics) Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data Analytics: Business Intelligence, Algorithms and Statistical Analysis (Predictive Analytics, Data Visualization, Data Analytics, Business Analytics, Decision Analysis, Big Data, Statistical Analysis) Data Analytics and Python Programming: 2 Bundle Manuscript: Beginners Guide to Learn Data Analytics, Predictive Analytics and Data Science with Python Programming Data Analytics For Beginners: Your Ultimate Guide To Learn and Master Data Analysis. Get Your Business Intelligence Right Ã¢â ¬â œ Accelerate Growth and Close More Sales (Data Analytics Book Series) Discovering Knowledge in Data: An Introduction to Data Mining (Wiley Series on Methods and Applications in Data Mining) Data Mining: Concepts and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems) Essential Oils: 50 Essential Oil Dog & Cat Recipes From My Essential Oil Private Collection: Proven Essential Oil Recipes That Work! (Essential Oil Pet Private Collection Book 1) Essential Oils: Essential Oil Recipe Book - 30 Proven

Essential Oil Recipes ::: My Essential Oil Private Collection Vol. 1 (Private Collection Essential Oils) Practical Data Management for Risk Data Aggregation and BCBS 239 Compliance Second Edition Data Mining: Practical Machine Learning Tools and Techniques (Morgan Kaufmann Series in Data Management Systems) Principles of Data Wrangling: Practical Techniques for Data Preparation