1. Record Nr. UNINA9910830622503321

Autore	Denis Daniel J. <1974->
Titolo	Applied univariate, bivariate, and multivariate statistics using Python : a beginner's guide to advanced data analysis / / Daniel J. Denis
Pubbl/distr/stampa	Hoboken, New Jersey : , : John Wiley & Sons, , [2021] ©2021
ISBN	1-5231-4320-7 1-119-57818-3 1-119-57820-5 1-119-57817-5
Descrizione fisica	1 online resource (300 pages)
Disciplina	519.5302855133
Soggetti	Statistics Multivariate analysis Python (Computer program language)
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di bibliografia	Includes bibliographical references and index.
Nota di contenuto	Cover -- Title Page -- Copyright Page -- Contents -- Preface -- 1. A Brief Introduction and Overview of Applied Statistics -- 1.1 How Statistical Inference Works -- 1.2 Statistics and Decision-Making -- 1.3 Quantifying Error Rates in Decision-Making: Type I and Type II Errors -- 1.4 Estimation of Parameters -- 1.5 Essential Philosophical Principles for Applied Statistics -- 1.6 Continuous vs. Discrete Variables -- 1.6.1 Continuity Is Not Always Clear-Cut -- 1.7 Using Abstract Systems to Describe Physical Phenomena: Understanding Numerical vs. Physical Differences -- 1.8 Data Analysis, Data Science, Machine Learning, Big Data -- 1.9 "Training" and "Testing" Models: What "Statistical Learning" Means in the Age of Machine Learning and Data Science -- 1.10 Where We Are Going From Here: How to Use This Book -- Review Exercises -- 2. Introduction to Python and the Field of Computational Statistics -- 2.1 The Importance of Specializing in Statistics and Research, Not Python: Advice for Prioritizing Your Hierarchy -- 2.2 How to Obtain Python -- 2.3 Python Packages -- 2.4 Installing a New Package in Python -- 2.5 Computing z-Scores in

Python -- 2.6 Building a Dataframe in Python: And Computing Some Statistical Functions -- 2.7 Importing a .txt or .csv File -- 2.8 Loading Data into Python -- 2.9 Creating Random Data in Python -- 2.10 Exploring Mathematics in Python -- 2.11 Linear and Matrix Algebra in Python: Mechanics of Statistical Analyses -- 2.11.1 Operations on Matrices -- 2.11.2 Eigenvalues and Eigenvectors -- Review Exercises -- 3. Visualization in Python: Introduction to Graphs and Plots -- 3.1 Aim for Simplicity and Clarity in Tables and Graphs: Complexity is for Fools! -- 3.2 State Population Change Data -- 3.3 What Do the Numbers Tell Us? Clues to Substantive Theory -- 3.4 The Scatterplot -- 3.5 Correlograms -- 3.6 Histograms and Bar Graphs.
3.7 Plotting Side-by-Side Histograms -- 3.8 Bubble Plots -- 3.9 Pie Plots -- 3.10 Heatmaps -- 3.11 Line Charts -- 3.12 Closing Thoughts -- Review Exercises -- 4. Simple Statistical Techniques for Univariate and Bivariate Analyses -- 4.1 Pearson Product-Moment Correlation -- 4.2 A Pearson Correlation Does Not (Necessarily) Imply Zero Relationship -- 4.3 Spearman's Rho -- 4.4 More General Comments on Correlation: Don't Let a Correlation Impress You Too Much! -- 4.5 Computing Correlation in Python -- 4.6 T-Tests for Comparing Means -- 4.7 Paired-Samples t-Test in Python -- 4.8 Binomial Test -- 4.9 The Chi-Squared Distribution and Goodness-of-Fit Test -- 4.10 Contingency Tables -- Review Exercises -- 5. Power, Effect Size, P-Values, and Estimating Required Sample Size Using Python -- 5.1 What Determines the Size of a P-Value? -- 5.2 How P-Values Are a Function of Sample Size -- 5.3 What is Effect Size? -- 5.4 Understanding Population Variability in the Context of Experimental Design -- 5.5 Where Does Power Fit into All of This? -- 5.6 Can You Have Too Much Power? Can a Sample Be Too Large? -- 5.7 Demonstrating Power Principles in Python: Estimating Power or Sample Size -- 5.8 Demonstrating the Influence of Effect Size -- 5.9 The Influence of Significance Levels on Statistical Power -- 5.10 What About Power and Hypothesis Testing in the Age of "Big Data"? -- 5.11 Concluding Comments on Power, Effect Size, and Significance Testing -- Review Exercises -- 6. Analysis of Variance -- 6.1 T-Tests for Means as a "Special Case" of ANOVA -- 6.2 Why Not Do Several t-Tests? -- 6.3 Understanding ANOVA Through an Example -- 6.4 Evaluating Assumptions in ANOVA -- 6.5 ANOVA in Python -- 6.6 Effect Size for Teacher -- 6.7 Post-Hoc Tests Following the ANOVA F-Test -- 6.8 A Myriad of Post-Hoc Tests -- 6.9 Factorial ANOVA -- 6.10 Statistical Interactions.
6.11 Interactions in the Sample Are a Virtual Guarantee: Interactions in the Population Are Not -- 6.12 Modeling the Interaction Term -- 6.13 Plotting Residuals -- 6.14 Randomized Block Designs and Repeated Measures -- 6.15 Nonparametric Alternatives -- 6.15.1 Revisiting What "Satisfying Assumptions" Means: A Brief Discussion and Suggestion of How to Approach the Decision Regarding Nonparametrics -- 6.15.2 Your Experience in the Area Counts -- 6.15.3 What If Assumptions Are Truly Violated? -- 6.15.4 Mann-Whitney U Test -- 6.15.5 Kruskal-Wallis Test as a Nonparametric Alternative to ANOVA -- Review Exercises -- 7. Simple and Multiple Linear Regression -- 7.1 Why Use Regression? -- 7.2 The Least-Squares Principle -- 7.3 Regression as a "New" Least-Squares Line -- 7.4 The Population Least-Squares Regression Line -- 7.5 How to Estimate Parameters in Regression -- 7.6 How to Assess Goodness of Fit? -- 7.7 R2 - Coefficient of Determination -- 7.8 Adjusted R2 -- 7.9 Regression in Python -- 7.10 Multiple Linear Regression -- 7.11 Defining the Multiple Regression Model -- 7.12 Model Specification Error -- 7.13 Multiple Regression in Python -- 7.14 Model-Building Strategies: Forward, Backward, Stepwise

-- 7.15 Computer-Intensive "Algorithmic" Approaches -- 7.16 Which Approach Should You Adopt? -- 7.17 Concluding Remarks and Further Directions: Polynomial Regression -- Review Exercises -- 8. Logistic Regression and the Generalized Linear Model -- 8.1 How Are Variables Best Measured? Are There Ideal Scales on Which a Construct Should Be Targeted? -- 8.2 The Generalized Linear Model -- 8.3 Logistic Regression for Binary Responses: A Special Subclass of the Generalized Linear Model -- 8.4 Logistic Regression in Python -- 8.5 Multiple Logistic Regression -- 8.5.1 A Model with Only Lag1 -- 8.6 Further Directions -- Review Exercises.
9. Multivariate Analysis of Variance (MANOVA) and Discriminant Analysis -- 9.1 Why Technically Most Univariate Models are Actually Multivariate -- 9.2 Should I Be Running a Multivariate Model? -- 9.3 The Discriminant Function -- 9.4 Multivariate Tests of Significance: Why They Are Different from the F-Ratio -- 9.4.1 Wilks' Lambda -- 9.4.2 Pillai's Trace -- 9.4.3 Roy's Largest Root -- 9.4.4 Lawley-Hotelling's Trace -- 9.5 Which Multivariate Test to Use? -- 9.6 Performing MANOVA in Python -- 9.7 Effect Size for MANOVA -- 9.8 Linear Discriminant Function Analysis -- 9.9 How Many Discriminant Functions Does One Require? -- 9.10 Discriminant Analysis in Python: Binary Response -- 9.11 Another Example of Discriminant Analysis: Polytomous Classification -- 9.12 Bird's Eye View of MANOVA, ANOVA, Discriminant Analysis, and Regression: A Partial Conceptual Unification -- 9.13 Models "Subsumed" Under the Canonical Correlation Framework -- Review Exercises -- 10. Principal Components Analysis -- 10.1 What Is Principal Components Analysis? -- 10.2 Principal Components as Eigen Decomposition -- 10.3 PCA on Correlation Matrix -- 10.4 Why Icebergs Are Not Good Analogies for PCA -- 10.5 PCA in Python -- 10.6 Loadings in PCA: Making Substantive Sense Out of an Abstract Mathematical Entity -- 10.7 Naming Components Using Loadings: A Few Issues -- 10.8 Principal Components Analysis on USA Arrests Data -- 10.9 Plotting the Components -- Review Exercises -- 11. Exploratory Factor Analysis -- 11.1 The Common Factor Analysis Model -- 11.2 Factor Analysis as a Reproduction of the Covariance Matrix -- 11.3 Observed vs. Latent Variables: Philosophical Considerations -- 11.4 So, Why is Factor Analysis Controversial? The Philosophical Pitfalls of Factor Analysis -- 11.5 Exploratory Factor Analysis in Python -- 11.6 Exploratory Factor Analysis on USA Arrests Data.
Review Exercises -- 12. Cluster Analysis -- 12.1 Cluster Analysis vs. ANOVA vs. Discriminant Analysis -- 12.2 How Cluster Analysis Defines "Proximity" -- 12.2.1 Euclidean Distance -- 12.3 K-Means Clustering Algorithm -- 12.4 To Standardize or Not? -- 12.5 Cluster Analysis in Python -- 12.6 Hierarchical Clustering -- 12.7 Hierarchical Clustering in Python -- Review Exercises -- References -- Index -- EULA.

Sommario/riassunto

"This book is an elementary beginner's introduction to applied statistics using Python. It for the most part assumes no prior knowledge of statistics or data analysis, though a prior introductory course is desirable. It can be appropriately used in a 16-week course in statistics or data analysis at the advanced undergraduate or beginning graduate level in fields such as psychology, sociology, biology, forestry, education, nursing, chemistry, business, law, and other areas where making sense of data is priority rather than formal theoretical statistics as one may have in a more specialized program in a statistics department. Mathematics used in the book is minimal, and where math is used, every effort has been made to unpack and explain it as clearly as possible. The goal of the book is to obtain results using software rather quickly, while at the same time not completely dismissing

important conceptual and theoretical features. After all, if you do not understand what the computer is producing, then the output will be quite meaningless. For deeper theoretical accounts, the reader is encouraged to consult other sources, such as the author's more theoretical book, now in its second edition (Denis, 2021), or a number of other books on univariate and multivariate analysis (e.g., Izenman, 2008; Johnson and Wichern, 2007). The book you hold in your hands is merely meant to get your foot in the door, and so long as that is understood from the outset, it will be of great use to the newcomer or beginner in statistics and computing. It is hoped that you leave the book with a feeling of having better understood simple to relatively advanced statistics, while also experiencing a little bit about what Python is all about. Python is used in performing and demonstrating data analyses throughout the book, however it should be emphasized that the book is not a specialty on Python itself. In this respect, the book does not contain a deep introduction to the software, nor does it go into the language that makes up Python computing to any significant degree. Rather, the book is much more "hands-on" in that code used is a starting point to generating useful results. That is, the code employed is that which worked for the problem under consideration, and which the user can amend or adjust afterward when performing additional analyses. When it comes to coding with Python, there are usually several ways of accomplishing similar goals. In places, we also cite code used by others, assigning proper credit. There already exist a plethora of Python texts and user manuals that feature the software in much greater depth. Those users wishing to learn Python from scratch and become specialists in the software and aspire to become an efficient and general-purpose programmer should consult those sources (e.g., see Guttag, 2013). For those who want some introductory exposure to Python on generating data-analytic results, and wish to understand what the software is producing, it is hoped that the current book will be of great use"--