Statistical Data Analytics. Foundations for Data Mining, Informatics, and Knowledge Discovery

Description:
A comprehensive introduction to statistical methods for data mining and knowledge discovery.

Applications of data mining and big data increasingly take center stage in our modern, knowledge-driven society, supported by advances in computing power, automated data acquisition, social media development and interactive, linkable internet software. This book presents a coherent, technical introduction to modern statistical learning and analytics, starting from the core foundations of statistics and probability. It includes an overview of probability and statistical distributions, basics of data manipulation and visualization, and the central components of standard statistical inferences. The majority of the text extends beyond these introductory topics, however, to supervised learning in linear regression, generalized linear models, and classification analytics. Finally, unsupervised learning via dimension reduction, cluster analysis, and market basket analysis are introduced.

Extensive examples using actual data (with sample R programming code) are provided, illustrating diverse informatic sources in genomics, biomedicine, ecological remote sensing, astronomy, socioeconomics, marketing, advertising and finance, among many others.

Statistical Data Analytics:
- Focuses on methods critically used in data mining and statistical informatics. Coherently describes the methods at an introductory level, with extensions to selected intermediate and advanced techniques.
- Provides informative, technical details for the highlighted methods.
- Employs the open-source R language as the computational vehicle along with its burgeoning collection of online packages to illustrate many of the analyses contained in the book.
- Concludes each chapter with a range of interesting and challenging homework exercises using actual data from a variety of informatic application areas.

This book will appeal as a classroom or training text to intermediate and advanced undergraduates, and to beginning graduate students, with sufficient background in calculus and matrix algebra. It will also serve as a source-book on the foundations of statistical informatics and data analytics to practitioners who regularly apply statistical learning to their modern data.

Contents:
Preface xiii
Part I Background: Introductory Statistical Analytics 1
1 Data analytics and data mining 3
1.1 Knowledge discovery: finding structure in data 3
1.2 Data quality versus data quantity 5
1.3 Statistical modeling versus statistical description 7
2 Basic probability and statistical description 10
2.1 Concepts in probability 10
2.1.1 Probability rules 11
2.1.2 Random variables and probability functions 12
2.1.3 Means, variances, and expected values 17
2.1.4 Median, quartiles, and quantiles 18
2.1.5 Bivariate expected values, covariance, and correlation 20
2.2 Multiple random variables 21
2.3 Univariate families of distributions 23
2.3.1 Binomial distribution 23
2.3.2 Poisson distribution 26
2.3.3 Geometric distribution 27
2.3.4 Negative binomial distribution 27
2.3.5 Discrete uniform distribution 28
2.3.6 Continuous uniform distribution 29
2.3.7 Exponential distribution 29
2.3.8 Gamma and chi–square distributions 30
2.3.9 Normal (Gaussian) distribution 32
2.3.10 Distributions derived from normal 37
2.3.11 The exponential family 41
3 Data manipulation 49
3.1 Random sampling 49
3.2 Data types 51
3.3 Data summarization 52
3.3.1 Means, medians, and central tendency 52
3.3.2 Summarizing variation 56
3.3.3 Summarizing (bivariate) correlation 59
3.4 Data diagnostics and data transformation 60
3.4.1 Outlier analysis 60
3.4.2 Entropy 62
3.4.3 Data transformation 64
3.5 Simple smoothing techniques 65
3.5.1 Binning 66
3.5.2 Moving averages 67
3.5.3 Exponential smoothing 69
4 Data visualization and statistical graphics 76
4.1 Univariate visualization 77
4.1.1 Strip charts and dot plots 77
4.1.2 Boxplots 79
4.1.3 Stem–and–leaf plots 81
4.1.4 Histograms and density estimators 83
4.1.5 Quantile plots 87
4.2 Bivariate and multivariate visualization 89
4.2.1 Pie charts and bar charts 90
4.2.2 Multiple boxplots and QQ plots 95
4.2.3 Scatterplots and bubble plots 98
4.2.4 Heatmaps 102
4.2.5 Time series plots 105
5 Statistical inference 115
5.1 Parameters and likelihood 115
5.2 Point estimation 117
5.2.1 Bias 118
5.2.2 The method of moments 118
5.2.3 Least squares/weighted least squares 119
5.2.4 Maximum likelihood 120
5.3 Interval estimation 123
5.3.1 Confidence intervals 123
5.3.2 Single–sample intervals for normal (Gaussian) parameters 124
5.3.3 Two–sample intervals for normal (Gaussian) parameters 128
5.3.4 Wald intervals and likelihood intervals 131
5.3.5 Delta method intervals 135
5.3.6 Bootstrap intervals 137
5.4 Testing hypotheses 138
5.4.1 Single–sample tests for normal (Gaussian) parameters 140
5.4.2 Two–sample tests for normal (Gaussian) parameters 142
5.4.3 Walds tests, likelihood ratio tests, and exact tests 145
5.5 Multiple inferences 148
Fax Order Form
To place an order via fax simply print this form, fill in the information below and fax the completed form to 646-607-1907 (from USA) or +353-1-481-1716 (from Rest of World). If you have any questions please visit http://www.researchandmarkets.com/contact/

Order Information
Please verify that the product information is correct.

Product Name: Statistical Data Analytics. Foundations for Data Mining, Informatics, and Knowledge Discovery
Web Address: http://www.researchandmarkets.com/reports/3025006/
Office Code: SCBRW5VB

Product Format
Please select the product format and quantity you require:

<table>
<thead>
<tr>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard Copy (Hard Back): USD 99 + USD 29 Shipping/Handling</td>
</tr>
</tbody>
</table>

* Shipping/Handling is only charged once per order.

Contact Information
Please enter all the information below in BLOCK CAPITALS

Title: [ ] Mr [ ] Mrs [ ] Dr [ ] Miss [ ] Ms [ ] Prof
First Name: ____________________________ Last Name: ____________________________
Email Address: * ____________________________
Job Title: ____________________________
Organisation: ____________________________
Address: ____________________________
City: ____________________________
Postal / Zip Code: ____________________________
Country: ____________________________
Phone Number: ____________________________
Fax Number: ____________________________

* Please refrain from using free email accounts when ordering (e.g. Yahoo, Hotmail, AOL)
Payment Information

Please indicate the payment method you would like to use by selecting the appropriate box.

☐ Pay by credit card: You will receive an email with a link to a secure webpage to enter your credit card details.

☐ Pay by check: Please post the check, accompanied by this form, to:

Research and Markets,
Guinness Center,
Taylors Lane,
Dublin 8,
Ireland.

☐ Pay by wire transfer: Please transfer funds to:

Account number 833 130 83
Sort code 98-53-30
Swift code ULSBIE2D
IBAN number IE78ULSB98533083313083
Bank Address Ulster Bank,
27-35 Main Street,
Blackrock,
Co. Dublin,
Ireland.

If you have a Marketing Code please enter it below:

Marketing Code: ___________________________________________

Please note that by ordering from Research and Markets you are agreeing to our Terms and Conditions at http://www.researchandmarkets.com/info/terms.asp

Please fax this form to:
(646) 607-1907 or (646) 964-6609 - From USA
+353-1-481-1716 or +353-1-653-1571 - From Rest of World