

Statistical Algorithms



Alpha
Science

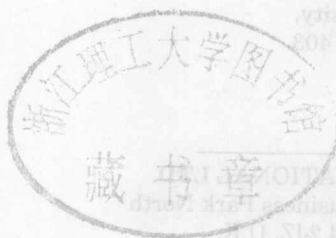
Rajan Chattamvelli

STATISTICAL ALGORITHMS



30807595

Rajan Chattamvelli



Alpha Science International Ltd.
Oxford, U.K.

Rajan Chattamvelli

Associate Professor of IT,
Periyar Maniammai University,
Thanjavur, Tamil Nadu 613 403
India

Copyright © 2012

ALPHA SCIENCE INTERNATIONAL LTD.
7200 The Quorum, Oxford Business Park North
Garsington Road, Oxford OX4 2JZ, U.K.

www.alphasci.com

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior written permission of the publisher.

Printed from the camera-ready copy provided by the Author.

ISBN 978-1-84265-727-0

Printed in India

Preface

This is an intermediate level textbook on statistics with emphasis on theory and algorithms. It integrates the multitude of algorithms available for common data analysis tasks. Numerical examples illustrate most of the algorithms. Pseudo code for the algorithms is given to benefit the students and researchers.

This book can be used for senior undergraduate or graduate level courses in statistical computing, statistical algorithms, computational methods in statistics, computer intensive methods, and engineering statistics. Some of the chapters can also be used for courses on mathematical statistics (ch.1-4,6-10,12), advanced statistical methods (ch.1, 5-12), actuarial statistics (ch.2-4,6,7), biostatistics, economics (ch.2-4,8), information technology (ch.1-5,12), numerical methods (ch.1,5,12), industrial statistics (ch.2-5,8-10,12), econometrics (ch.1-7,9-11), management(ch.2-4, 8-10), medical statistics, commerce (ch.2-4,6) and computer science (ch.1-5,8,11,12). This is also an ideal book for researchers in various fields. The study of statistical algorithms is also important for scientists and professionals. A brief overview of the chapters is given below.

A general discussion of algorithms appears in Chapter 1. Complexity of algorithms is clearly discussed and types of complexities are described. Complexity notations are briefly reviewed in §1.2 (1-3). A recursive algorithm for sample mean appears in page 1-3. A summary table of complexity of algorithms can be found in page 1-5. Algorithm design techniques are discussed next. Several examples of recursive and iterative algorithms are given. Efficiency of algorithms is discussed in §1.4(p.1-14), followed by data dependency on algorithms. Algorithms for minimum and maximum (§1.5), powers, combinations, k^{th} smallest or largest element in an unsorted list (§1.7) are discussed.

Chapter 2 introduces popular measures of central tendency. Topics covered include arithmetic, geometric, harmonic and quadratic means, median and mode. Means for frequency data and weighted means are also discussed. Several novel algorithms to update these measures when data arrive afresh are given (p.2-5). Algorithms for trimmed-mean (page 2-7), sample median (page 2-12), mode (page 2-14) etc are discussed.

Chapter 3 introduces dispersion measures. Sample range (p.3-1), inter-quartile range, average absolute deviation (p.3-4), variance and standard deviation (p.3-5), etc are discussed. This is followed by a discussion of outliers and their detection (p.3-9). The chapter ends with a section on data transformations (p.3-11).

The chapter 4 discusses various measures of skewness and kurtosis. Absolute and relative measures are discussed in p.4-2. Popular skewness measures like Pearson's measure(p.4-3), Galton's measure (p.4-4), Kelly's measure and Bowley's measure (p.4-5) are introduced, and a new measure is proposed (p.4-5). This is followed by a discussion of kurtosis measures (p.4-8). The chapter ends with a review of L-kurtosis (p.4-10).

Chapter 5 is on matrix algorithms. Several statistical procedures like multiple regression, principal component analysis etc use matrix algorithms. This chapter starts with a general discussion of matrices. These are grouped under various categories for ease of comprehension. Quadratic forms are briefly introduced in section §5.5. Some of the popular matrix algorithms are introduced and illustrated with numerical examples. These include matrix inverse (p.5-10), matrix decompositions (p.5-15), eigen values (p.5-20), iterative matrix algorithms (p.5-28), etc. Linear equation algorithms are illustrated in §5.9.

Chapter 6 on Simple Linear Regression (SLR) gives algorithms for fitting linear regression model. Some variants of SLR are also discussed. Multiple Linear Regression (MLR) is described in §6.3. Multi-collinearity and its detection techniques are introduced in §6.4 and §6.5. This is followed by outlier detection (§6.6), variable selection methods (§6.7), and partial least squares (§6.8).

Chapter 7 gives a thorough discussion of measures of association. Sample covariance and correlation are introduced in §7.2 and §7.3. This is followed by other types of correlation like rank-correlation, Spearman's ρ , Kendall's τ , point biserial correlation, and ϕ -coefficient.

Chapter 8 is on distance-based methods. These include clustering, nearest neighbor algorithms, etc. Distance metrics that are extensively used in nearest neighbor algorithms, clustering algorithms, etc are discussed in the first few sections (p. 8-3 to 8-6). This is followed by similarity metrics. Popular clustering algorithms are discussed next. The chapter ends with a discussion of Nearest Neighbour Algorithms (NNA).

Chapter 9 discusses important discrete distributions. Some simple algorithms for the most popular distributions like Bernoulli, binomial, Poisson, geometric, hypergeometric, negative binomial and discrete uniform distributions are discussed in this chapter. Important properties are briefly summarised, and used in chapter 12 on random numbers.

The tenth chapter is on continuous univariate distributions. It discusses exponential, normal, gamma, beta, Cauchy, chi-square, t and F

distributions. A new family of distributions is introduced in §10.4 (p.10-4). Other distributions like skew-normal, log-normal, inverse Gaussian and arc-sin are briefly mentioned. Their properties and tail areas are summarised and used in chapter 12.

Chapter 11 is on EM Algorithm (EMA). This is a popular alternative to estimate parameters of a statistical population in the presence of missing data. A discussion on convexity of functions can be found in the beginning of this chapter. This is not needed from a practitioners view point, but is used in the derivation of the EM algorithm. The EMA is derived in page 11-8. Several examples of its use appear next. Variants of EM algorithms are also briefly reviewed.

The last chapter is on generating random numbers for simulation studies. Topics covered include random shuffling, congruential generators (linear and additive), functional composition method, etc. Generating random numbers for most of the important discrete and continuous distributions are discussed, and pseudocode for some of the algorithms are given. The last section discusses correlated random numbers. Chapters 9, 10 and 12 can be used for a course on simulation.

Chapter 2, 3, 7, 9 and 10 can be used for undergraduate courses in statistics, mathematics, econometrics, and business management. Other chapters may be used as 'supplementary reading' for various courses mentioned above. Professionals and practitioners in various fields can also use the book for self-study.

A large number of useful code-snippets and pseudo-code of algorithms appear throughout the book¹. Educational use of this code is permitted, provided that it is duly acknowledged. Every software that uses the code in this book must document it as "Source: RAJAN CHATTAMVELLI (STATISTICAL ALGORITHMS)" without fail. Please contact the author at the following email for commercial use, or for porting to specific languages. Any misuse must be reported to the author. The author or publisher is not liable for any damage, loss, malfunction or hazard resulting from the use of the code.

A carefully selected set of exercises has been provided to benefit the students and self-study professionals. Answers and hints are provided for selected exercises. Any suggestions and comments for improvement are welcome. All suggestions should be sent to dmmbook@gmail.com. Up-to-date errata will be made available upon request.

Rajan Chattamvelli
Tanjore, Tamil Nadu.

¹The semicolons that terminate each statement (in C/C++,Java) are omitted in all pseudocodes given in subsequent chapters to make it easy to port to VB, FORTRAN, and other programming languages without statement terminators.

Table of Contents

1. INTRODUCTION TO ALGORITHMS	1-1
1.1 Introduction	1-1
1.1.1 Categorisations of Algorithms	1-2
1.2 Complexity Notations	1-3
1.2.1 Importance of Algorithms	1-6
1.3 Algorithm Design	1-7
1.3.1 Algorithmic Paradigms	1-8
1.4 Efficiency of Algorithms	1-14
1.4.1 Algorithm vs Data	1-17
1.5 Algorithms For Minimum and Maximum	1-18
1.6 Algorithm to find Middle of Three Elements	1-19
1.7 Algorithm to find k^{th} Smallest or Largest Member	1-19
1.8 Exercises	1-20
References	1-24
2. MEASURES OF CENTRAL TENDENCY	2-1
2.1 Introduction	2-1
2.1.1 Population vs Sample	2-1
2.1.2 Parameter vs Statistic	2-2
2.2 Measures of Location	2-3
2.3 Arithmetic Mean	2-3
2.3.1 Updating Formula For Sample Mean	2-6
2.3.2 Short-cut Method for Mean	2-8
2.3.3 Weighted Mean	2-8
2.3.4 Updating Formula For Weighted Sample Mean	2-9
Advantages of Mean	2-9
2.4 Median	2-10
Advantages of Median	2-13
2.5 Mode	2-13
2.5.1 Advantages of Mode	2-14
2.6 Geometric Mean	2-16
2.7 Harmonic Mean	2-17
2.8 Which Mean is Most Appropriate?	2-18
2.9 Quartiles and Percentiles	2-19
2.9.1 Five-number Summary	2-19
2.10 Exercises	2-19

References	2-22
3. MEASURES OF DISPERSION	3-1
3.1 Dispersion	3-1
3.1.1 Range	3-2
3.1.2 Inter-Quartile Range	3-3
3.1.3 Averaged Absolute Deviation (AAD)	3-5
3.1.4 Variance Ratio	3-6
3.1.5 Variance	3-6
3.1.6 Recursive Algorithm For Sample Variance	3-9
3.2 Outliers in Data	3-10
3.2.1 Spatial vs Temporal Outliers	3-10
3.2.2 Graphical Detection of Outliers	3-11
3.3 Data Transformations	3-12
3.3.1 Change of Origin	3-12
3.3.2 Change of Scale	3-13
3.3.3 Standard Normalisation	3-13
3.3.4 Nonlinear Transformations	3-14
3.4 Exercises	3-14
References	3-16
4. MEASURES OF SKEWNESS AND KURTOSIS	4-1
4.1 Skewness	4-1
4.2 Measures of Skewness	4-2
4.2.1 Absolute and Relative Measure of Skewness	4-3
4.2.2 Pearson's Measure of Skewness	4-3
4.2.3 Galton's Measure of Skewness	4-4
4.2.4 Bowley's Measure of Skewness	4-5
4.2.5 New Measures of Skewness	4-5
4.3 Kurtosis	4-8
4.3.1 Measures of Kurtosis	4-9
4.3.2 Pearson's Kurtosis Measure	4-9
4.3.3 L-kurtosis	4-10
4.4 Exercises	4-10
References	4-13
5. MATRIX ALGORITHMS IN STATISTICS	5-1
5.1 Introduction	5-1
5.2 Vectors	5-2
5.3 Matrices	5-3
5.3.1 Special Vectors and Matrices	5-4
Special Vectors	5-4
Special Matrices	5-4
5.3.2 Determinants	5-7
5.3.3 Rules of Determinant	5-8
5.3.4 Matrix Inverse	5-8
5.4 Some Matrix Rules	5-9
5.4.1 Rules of Inverse	5-9

5.5	Quadratic Forms (QF)	5-11
5.6	Matrix Algorithms	5-11
5.6.1	Maximum and Minimum Elements of a Matrix	5-12
5.6.2	Matrix Multiplication Algorithms	5-13
5.6.3	Naive Matrix Multiplication	5-13
5.6.4	Partitioned Matrices	5-14
5.6.5	Strassen's Matrix Multiplication	5-14
5.7	Decomposition Algorithms	5-15
5.7.1	LU-Decomposition Algorithm (LU-DA)	5-15
5.7.2	QR-Decomposition Algorithm (QR-DA)	5-17
5.7.3	Cholesky-Decomposition	5-18
5.8	Eigenvalues and Eigenvectors	5-20
5.8.1	Properties of Eigenvalues	5-20
5.8.2	Eigenvalue Decomposition	5-21
5.8.3	Largest Eigenvalue	5-23
5.8.4	SVD-decomposition	5-23
5.9	Generalised Inverse of Matrices (GIM)	5-23
5.10	Solving Linear Equations	5-24
5.10.1	Linear Equation	5-24
5.10.2	Gaussian Elimination Algorithm (GEA)	5-25
5.10.3	Gaussian Elimination for Tri-diagonal Systems	5-27
5.11	Inhomogeneous Linear Equations	5-27
5.12	Iterative Algorithms for Linear Equations	5-27
5.12.1	Jacobi's Iterative Method (JIM)	5-28
5.12.2	Gauss-Siedel Iterative Method (G-SIM)	5-29
5.12.3	Successive Over-Relaxation Method (SORM)	5-30
5.13	Exercises	5-30
	References	5-34
6.	LINEAR REGRESSION	6-1
6.1	Regression Basics	6-1
6.1.1	Scatterplots and Regression	6-1
	Advantages of Scatter Plots	6-2
6.2	Simple Linear Regression	6-3
6.2.1	Ordinary Least Squares (OLS)	6-4
6.2.2	Centered Data	6-6
6.3	Multiple Linear Regression (MLR)	6-8
6.4	Multicollinearity in Regression	6-9
6.5	Multicollinearity Detection Techniques	6-10
6.5.1	Correlation-Matrix Based Method	6-10
6.5.2	VIF Method	6-10
6.5.3	Eigen Values of Dispersion Matrix	6-10
6.6	Outlier Detection in MLR	6-11
6.7	Variable Selection Methods	6-11
6.7.1	Forward selection method (FSM)	6-13
6.7.2	Backward Elimination Method (BEM)	6-13
6.7.3	Stepwise Regression Method (SRM)	6-15
	The R^2 Method	6-15

	Mallow's C_p Method	6-15
6.8	Nonlinear Regression	6-16
6.9	Partial Least Squares	6-16
6.10	Exercises	6-17
	References	6-18
7.	MEASURES OF ASSOCIATION	7-1
7.1	Introduction	7-1
7.2	Sample Covariance	7-2
7.2.1	Recursive Algorithm For Covariance	7-6
7.3	From Scatterplot to Correlation	7-8
	Geometric Interpretation	7-10
	Properties of Correlation Coefficient	7-11
7.3.1	Spearman's Rank Correlation Coefficient	7-12
	Advantages of Spearmans ρ	7-13
	Duplicate Rank Problem	7-14
	Kendall's τ	7-14
7.3.2	Point Biserial Correlation Coefficient	7-14
7.4	Contingency Tables	7-14
7.4.1	Phi-Coefficient	7-15
7.5	Exercises	7-16
	References	7-18
8.	DISTANCE BASED ALGORITHMS	8-1
8.1	Meaning of Distance-Based Algorithms	8-1
8.2	Cluster Analysis	8-1
8.2.1	Cluster Analysis Step-by-Step	8-2
8.3	Distance Metrics	8-3
8.3.1	Desirable Properties of a Metric	8-3
8.3.2	Euclidean Distance (ED) Metric (L_2 norm)	8-3
8.3.3	Manhattan Metric (L_1 Norm)	8-4
8.3.4	Minkowski Metric (L_p Norm)	8-5
8.3.5	Mahalanobis' Distance Metric	8-6
8.3.6	Chebychev Metric (L_∞ metric)	8-6
8.4	Similarity Metrics	8-7
8.4.1	Dice Metric	8-7
8.4.2	Jaccard Metric	8-7
8.5	Clustering Algorithms	8-8
8.6	Hierarchical Clustering Algorithms (HCA)	8-8
8.6.1	Agglomerative Algorithm	8-9
8.6.2	Divisive Algorithm	8-11
8.7	Partitioning Algorithms	8-13
8.7.1	K-means Clustering Algorithm	8-14
8.7.2	Drawbacks of The K-means Algorithm	8-16
8.7.3	Density-based Methods	8-18
8.8	Cluster Validation Techniques (CVT)	8-18
8.9	Nearest Neighbour Algorithms (NNA)	8-19
8.10	Software for Clustering and K-NN	8-20

8.11 Exercises	8-21
References	8-23

9. DISCRETE DISTRIBUTIONS 9-1

9.1 Introduction	9-1
9.2 Discrete Uniform Distribution	9-1
9.2.1 Properties of Discrete Uniform Distribution	9-2
9.3 Bernoulli Distribution	9-2
9.3.1 Properties of Bernoulli Distribution	9-2
9.4 Binomial Distribution	9-2
9.4.1 Properties of Binomial Distribution	9-3
9.4.2 Tail Probabilities	9-4
9.5 Multinomial Distribution	9-5
9.5.1 Properties of Multinomial Distribution	9-6
9.6 Geometric Distribution	9-6
9.6.1 Properties of Geometric Distribution	9-7
9.6.2 Tail Probabilities	9-8
9.7 Hypergeometric Distribution (HGD)	9-8
9.7.1 Properties of Hypergeometric Distribution	9-9
9.7.2 Tail Probabilities	9-9
9.8 Negative Hypergeometric Distribution	9-10
9.9 Negative Binomial Distribution	9-10
9.9.1 Properties of Negative Binomial Distribution	9-11
9.9.2 Tail Probabilities	9-11
9.10 Poisson Distribution	9-11
9.10.1 Properties of Poisson Distribution	9-12
9.10.2 Tail Probabilities	9-13
9.11 Logarithmic Distribution	9-14
9.11.1 Properties of Logarithmic Distribution	9-14
9.12 Exercises	9-14
References	9-17

10. CONTINUOUS DISTRIBUTIONS 10-1

10.1 Introduction	10-1
10.2 Continuous Uniform Distribution	10-1
10.2.1 Properties of Continuous Uniform Distribution	10-2
10.3 Exponential Distributions	10-2
10.3.1 Properties of Exponential Distribution	10-2
10.3.2 Tail Areas	10-3
10.3.3 Double Exponential Distribution	10-3
Properties of Double Exponential Distribution	10-3
10.4 Bernoulli-Erlangian Distribution	10-4
10.5 Gamma Distribution	10-4
10.5.1 Properties of Gamma Distribution	10-5
10.5.2 Tail Areas	10-5
10.6 χ^2 Distribution	10-6
10.6.1 Properties of χ^2 Distribution	10-6
10.6.2 Tail Areas	10-7

10.7	Beta Distribution	10-7
10.7.1	Properties of Beta Distribution	10-7
10.7.2	Tail Areas	10-8
10.8	Normal Distribution	10-9
10.8.1	Properties of Normal Distribution	10-10
10.8.2	Tail Areas	10-10
10.8.3	Skew-normal Distribution	10-11
10.8.4	Lognormal Distribution	10-11
10.9	Cauchy Distribution	10-11
	Properties of Cauchy Distribution	10-12
10.10	Student's T Distribution	10-12
	Properties of Student's t Distribution.	10-12
	Tail Areas	10-13
10.11	F Distribution	10-14
	Properties of F Distribution	10-15
	Tail Areas	10-15
10.12	Inverse Gaussian Distribution (IGD)	10-15
	Properties of IGD	10-15
10.13	Arc-Sine Distribution	10-16
	Properties of Arc-Sine Distribution	10-16
10.14	Exercises	10-16
	References	10-19
11.	THE EM ALGORITHM	11-1
11.1	Introduction	11-1
11.1.1	Convex Functions	11-2
11.1.2	Maximum Likelihood Estimation (MLE)	11-4
	Finite Mixture Models (FMM)	11-6
11.1.3	Structure of EM Algorithm	11-6
11.2	Derivation of EM Algorithm	11-7
11.2.1	Censored Data Analysis	11-10
11.2.2	Convergence Analysis of EM algorithm.	11-14
11.3	Variants of EM algorithm	11-16
11.3.1	K-Means Algorithm as EM Algorithm.	11-16
11.4	Exercises	11-17
	References	11-19
12.	RANDOM NUMBERS	12-1
12.1	Introduction	12-1
12.2	Random Number Generators	12-1
12.2.1	Random Shuffling	12-3
12.2.2	Linear Congruential Generators (LCG)	12-3
12.2.3	Additive Congruential Generators (ACG)	12-5
12.2.4	Functional Composition Generators (FCG)	12-5
12.3	Random Numbers From Distributions	12-5
12.3.1	Inverse Distribution Function Method	12-6
12.3.2	Range Splitting Method	12-8
12.3.3	Finite Sum Method	12-8

12.3.4	Accept-Reject Method	12-9
12.3.5	Simple Transformation Method	12-10
12.3.6	Using Central Limit Theorem (CLT)	12-13
12.4	Correlated Random Numbers	12-14
12.5	Monte Carlo Methods (MCM)	12-14
12.5.1	MCMC methods	12-15
12.6	Exercises	12-15
	References	12-17
Appendix	– Solutions to Selected Exercises	A-1
	List of Acronyms	A-10
	List of Internet Tables	A-11
	List of Journals and Resources	A-12
Author Index		AI-1
Subject Index		SI-1

List of Figures

1.1	Growth of linear and exponential functions	1-12
1.2	Recursive Combinations evaluation tree	1-15
2.1	Mean, median or mode?	2-6
2.2	Measures of Central Tendency	2-10
3.1	Variance – Small and Large	3-7
4.1	Positive vs Negative Skew Distributions	4-2
6.1	Customer age vs purchase amount of vitamin supplements	6-9
8.1	Dendrogram (single linkage) for data in table 8.1.	8-13
8.2	Scatterplot of data in table 8.1.	8-13
9.1	Two Binomial Distributions $B(20,.2)$ and $B(20,.5)$	9-3
9.2	Skewed and Symmetric Binomials	9-5
9.3	Two Geometric Distributions	9-7
9.4	Two Negative Binomial Distributions	9-10
9.5	Three Poisson Distributions	9-12
10.1	Exponential Distributions for $\lambda=2, 4, 12$	10-3
10.2	Scaled Bernoulli-Gaussian Distribution	10-4
10.3	Bernoulli-Erlangian Distribution for $p=.2, .9, .99$	10-4
10.4	Gamma Distribution	10-5
10.5	Chi-squared Distribution.	10-6
10.6	Comparison of Standard Normal and Student's T Distri- butions with 5 and 13 degrees of freedom.	10-10
12.1	The CDF method	12-3
12.2	The Envelop method	12-3

List of Tables

1.1	Common Complexity Classes with Examples	1-5
2.1	Total money spent on Ice-cream (per week) & total hours TV Viewed (per week) by school kids	2-11
2.2	Predicting sales price of used cars	2-20
3.1	Marks obtained by students in an exam	3-4
3.2	Cumulative marks obtained by students	3-5
3.3	Average absolute deviation of student marks	3-6
3.4	Fuel filling by vehicles	3-8
3.5	Fuel filling example – mean and variance	3-9
3.6	Insurance premiums & # of Chat requests received	3-15
3.7	Odometer reading and age of cars	3-16
4.1	Skewness test for marks obtained by students	4-7
4.2	A comparison of skewness measures	4-8
4.3	Age of cars and servicing costs	4-11
5.1	Storing Special Matrices as One-dimensional Array	5-12
5.2	Properties of Matrix Eigenvalues and Eigenvectors	5-22
5.3	Progress of Convergence of Jacobi's Algorithm	5-29
6.1	Computing Table for Simple Linear Regression	6-4
6.2	Computing Table for Simple Ordinal Linear Regression . .	6-7
6.3	Multicollinearity Detection Techniques	6-10
6.4	Average number of minutes late per week	6-17
6.5	Average monthly electricity bill vs number of rooms . . .	6-18
7.1	Height and weight of 10 adult persons	7-2
7.2	Height (scaled) and weight of 10 adult persons	7-3
7.3	Average market price of a brand of used cars (in '00) . . .	7-5
7.4	Variance of Height (scaled) and Weight of 10 Adults . . .	7-9
7.5	Family income vs annual travel spending	7-9
7.6	Computing Table for Sample Correlation	7-10
7.7	Customer satisfaction index for regular visitors in a super- market	7-13
7.8	Auto-accident break-up among left & right handed drivers	7-15
7.9	Web-site ratings by 8 users	7-17
8.1	Income and insurance premium amount of 25 customers . .	8-10
8.2	Results of clustering using k-means (k=3) and Hierarchi- cal clustering (with complete linkage, Ward's method) . .	8-17
8.3	Monthly income and total minutes online per week	8-22
11.1	Table for Estimating Mixture Model Parameters	11-16

11.2	Table for Exercise 19	11-18
12.1	Inversion Method for Some Statistical Distributions	12-8
12.2	Random Numbers from Some Statistical Distributions	12-9
1	List of Acronyms	A-10
2	List of Internet Tables	A-11
3	List of Related Journals and Resources	A-12

1.1	Common Complex Classes with Examples	1-5
2.1	Total money spent on ice-cream (per week) & total hours	2-11
2.2	TV Viewed (per week) by school kids	2-20
3.1	Predicting sales price of used cars	3-4
3.1	Marks obtained by students in an exam	3-5
3.2	Cumulative marks obtained by students	3-6
3.3	Average absolute deviation of student marks	3-8
3.4	Fuel filling by vehicles	3-9
3.5	Fuel filling example - mean and variance	3-15
3.6	Insurance premiums & % of Car requests received	3-16
3.7	Owner's reading and age of cars	4-7
4.1	Shewes test for marks obtained by students	4-8
4.2	A comparison of students' measures	4-11
4.3	Age of cars and servicing costs	5-12
5.1	Storing Special Matrices as One-dimensional Array	5-22
5.2	Properties of Matrix Eigenvalues and Eigenvectors	5-23
5.3	Progress of Convergence of Jacobi's Algorithm	5-29
6.1	Computing Table for Simple Linear Regression	6-4
6.2	Computing Table for Simple Ordinal Linear Regression	6-7
6.3	Multicollinearity Detection Techniques	6-10
6.4	Average number of minutes late per week	6-17
6.5	Average monthly electricity bill vs number of rooms	6-18
7.1	Height and weight of 10 adult persons	7-2
7.2	Height (scaled) and weight of 10 adult persons	7-3
7.3	Average market price of a brand of used cars (in '00)	7-5
7.4	Variance of Height (scaled) and Weight of 10 Adults	7-9
7.5	Family income vs annual travel spending	7-9
7.6	Computing Table for Sample Correlation	7-10
7.7	Customer satisfaction index for regular visitors in a super-market	7-13
7.8	Auto-accident break-up among left & right handed drivers	7-15
7.9	Website ratings by 8 users	7-17
8.1	Income and insurance premium amount of 25 customers	8-10
8.2	Results of clustering using k-means (k=3) and Hierarchical clustering (with complete linkage, Ward's method)	8-17
8.3	Monthly income and total minutes online per week	8-22
11.1	Table for Estimating Mixture Model Parameters	11-16

List of Algorithms

1.1	Recursive Algorithm for the Sample Mean	1-3
1.2	Iterative Algorithm for the Sample Mean	1-4
1.3	Two Iterative Algorithms for Powers (x^n)	1-7
1.4	Two Recursive Algorithms for x^n for any n (+ve or -ve)	1-8
1.5	Dynamic Programming Algorithm of Combinations	1-11
1.6	Pseudocode for Iterative & Recursive Combinations	1-13
1.7	Algorithm to find Maximum and Minimum	1-14
1.8	Approx. Factorial of +ve integer directly & using log	1-16
1.9	Iterative Logarithmic and Recursive algorithms to find Combinations of n things taken x at a time ($x \leq n$)	1-17
1.10	Iterative algorithm to find the Minimum and Maximum	1-18
1.11	Middle and MinMax of 3 Elements	1-19
1.12	Algorithm to find k-th smallest or k-th largest member	1-20
1.13	Pseudocode for question 19	1-22
1.14	Pseudocode for question 37	1-24
2.1	Algorithm for Updating Sample Mean (add/delete data)	2-5
2.2	Algorithm for 1-Trimmed Mean	2-7
2.3	Recursive Algorithm for Weighted Sample Mean	2-8
2.4	Algorithm for the Sample Median	2-12
2.5	Algorithm for the Sample Mode	2-14
2.6	Algorithm for Sample Geometric Mean	2-16
2.7	Recursive Algorithm for Sample Geometric Mean	2-17
2.8	Algorithm for Sample Harmonic Mean	2-18
2.9	Iterative Error-Corrected Algorithm for Sample Mean	2-22
3.1	Algorithm for Sample Range	3-2
3.2	Algorithm for 1-Trimmed Range	3-3
3.3	Recursive Algorithm for Sample Range	3-4
3.4	Algorithm for Sample Mean Absolute Deviation	3-7
3.5	Recursive Algorithms for the Sample Variance	3-10
4.1	Algorithm for Pearson's η	4-4
5.1	Algorithm for Matrix Inverse	5-10
5.2	Algorithm for Minimum and Maximum Matrix entries	5-13
5.3	Naive and Read-optimised Matrix Multiplication	5-14
5.4	Iterative Algorithm for LU- and QR Decomposition	5-16
5.5	Cholesky-Algorithm for Matrix-decomposition	5-18
5.6	Givens Rotation for matrix-decomposition	5-19
5.7	Power method for the largest eigen value	5-21